

ELMÉLETI KIEGÉSZÍTÉSEK A 2. FEJEZETHEZ

1. Közölt és valódi osztályköz-határok

A 2.2. alfejezetben az osztályközök határaiként olyan Y_{i0} , Y_{i1} osztópontok – osztályköz-határok – választását javasoltuk, melyek egymást át nem fedő osztályközöket jelölnek ki. Ehhez az Y_i alapadatokkal *azonos pontosságú* határokat kell választani, és arra is ügyelni kell, hogy egyik osztályköz felső határa sem egyezhet meg a következő osztályköz alsó határával. Az e két feltételnek eleget tevő határokat gyakran *közölt* osztályköz-határoknak nevezik. Minden olyan esetben, amikor az alapadatok valamilyen *adott pontosságra kerekített* értékek, mód van ún. *valódi határok* megadására is. Ha az alapadatok legutolsó szignifikáns számjegye 10^{sz} helyértékű, akkor a valódi határok az

$$\tilde{Y}_{i0} = Y_{i0} - \frac{1}{2} \cdot 10^{sz}, \quad \tilde{Y}_{i1} = Y_{i1} + \frac{1}{2} \cdot 10^{sz} \quad (1)$$

módon határozhatók meg a közölt határokból kiindulva. Az $\tilde{Y}_{i0}$ és $\tilde{Y}_{i1}$ által határolt osztályközök már *hézagmentesen* illeszkednek egymáshoz. Az alapadatoknak az egyes osztályközökbe való egyértelmű besorolhatóságát ilyenkor úgy lehet biztosítani, hogy az i -edik osztályköz határait az

$$\tilde{Y}_{i0} \leq Y_i < \tilde{Y}_{i1} \quad (2)$$

módon értjük.

Ha a 288 magyar város osztályközös gyakorisági sorba rendezését nem a 2.3. táblázatban látható módon végeznénk, hanem a városok *ezer főre kerekített* népességszáma alapján, akkor ehhez közölt és valódi határokat egyaránt lehetne használni. Az alábbi táblázat a lehetséges közölt határokat és az azoknak megfelelő valódi határokat tartalmazza. A valódi határok (1) felhasználásával kerültek meghatározásra, tekintetbe véve azt, hogy az ezer főre kerekített adatok abszolút hibakorlátja $\pm 0,5$ ezer fő.

A népesség száma (ezer fő)	
közölt	valódi
határok	
1 – 5	0,5 – 5,5
6 – 10	5,5 – 10,5
11 – 20	10,5 – 20,5
21 – 40	20,5 – 40,5
41 – 70	40,5 – 70,5
71 – 110	70,5 – 110,5
111 – 160	110,5 – 160,5
161 – 210	160,5 – 210,5

A valódi határok a közölt határokkal szemben azoknak az intervallumoknak a határait adják meg, melyekbe eső értékek a szokásos kerekítési szabály alkalmazása után a megfelelő közölt határok közé esnek.

Az osztályközepek nagysága nem függ attól, hogy a közölt vagy a valódi határok alapján kerülnek meghatározásra, de az osztályközök hossza nem független attól. Elvileg a kvantilisek és a módusz osztályközös gyakorisági sorból történő becsléséhez valódi határokat kell használni. A gyakorlatban azonban sokszor a közölt határokat használják. Ha azonban a közölt határok esetében élünk 2.2. alfejezetben leírt kerekítéssel, akkor az éppen a valódi határok használatával egyenértékű.

2. A kvantilisek általános definíciója

Ahhoz, hogy a kvantilisekkel határolt osztályközökben a gyakoriságok és relatív gyakoriságok *pontosan egyformák* legyenek, teljesülnie kell a következő két feltételnek:

- a megfigyelt sokaság nagysága (N) k -nak egész számú többszöröse kell hogy legyen;
- a kvantilis értékek nem eshetnek egybe egyik ténylegesen előforduló Y_i ismérvértékkel sem.

Ha a megfigyelések száma k -nak nem egész számú többszöröse, és/vagy az eddig ismertetett módon meghatározott kvantilisek valamelyike egybeesik egy ténylegesen is fellépő ismérvértékkel, akkor az nem tehet pontosan eleget a kvantilisekre korábban adott definíciónak. Ennek esélye különösen akkor nagy, ha a megfigyelések száma k -hoz képest kicsi, és/vagy az alapadatok között sok egymással egyező érték van. Ezért ilyenkor a kvantilisek számítása és használata kerülendő. Ez még a medián esetében is így van.

A teljesség kedvéért megjegyezzük, hogy a kvantilisek ilyen esetekben is meghatározhatók, de ekkor definíciójuk, illetve értelmezésük a következő:

A p -ed rendű Y_p kvantilis az a szám, aminél az összes előforduló ismérvérték legalább p -ed része nem nagyobb és legalább $(1-p)$ -ed része nem kisebb; ahol $0 < p < 1$. Az összes előforduló ismérvérték közé maga az Y_p kvantilis érték is beleértendő, akár szerepel a ténylegesen előforduló ismérvértékek között, akár nem.

Ez a definíció már egyértelműen meghatározhatóvá teszi a kvantiliseket akkor is, ha a fent megfogalmazott két feltétel nem teljesül. Ha N k -hoz képest nagy, és/vagy az Y_i alapadatok között csak kevés egymással megegyező érték van, akkor a korábbi és a mostani definíció használata szinte pontosan azonos kvantiliseket eredményez. A kvantilisek meghatározására szolgáló, 2.3. alfejezetben ismertetett módszerek valójában a kvantilisek most megfogalmazott általános definícióján alapszanak. Ha a kvantilisek közelítő értékét osztályközös gyakorisági sorból kiindulva határozzuk meg, akkor ehhez elvileg a megfelelő osztályköz valódi határai alapján meghatározott osztályköz-hosszúságot kellene használni. A gyakorlatban azonban általában nem így járnak el, de ez legtöbbször alig befolyásolja a keresett kvantilis közelítő értékét.

3. A szórásnégyzet és az átlagok összefüggése

A szórás, illetve szórásnégyzet (2-22) és (2-23) formulája számolásra nem nagyon alkalmas. A két formula azonban a

$$\sum d_i^2 = \sum Y_i^2 - N\bar{Y}^2, \text{ illetve } \sum f_i d_i^2 = \sum f_i Y_i^2 - N\bar{Y}^2 \quad (3)$$

összefüggések felhasználásával a

$$\sigma^2 = \bar{Y}_q^2 - \bar{Y}^2 \quad (4)$$

alakra hozható. A (3) összefüggések használata már nem teszi szükségessé az átlagtól vett d_i eltérések egyenkénti kiszámítását, ami gyakran körülményes. Különösen akkor, ha az átlag nem kerek szám.

A *súlyozatlan* esetben a (3) összefüggés az alábbi módon igazolható:

$$\sum d_i^2 = \sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - 2\bar{Y} \sum Y_i + \sum \bar{Y}^2 = \sum Y_i^2 - 2N\bar{Y}^2 + N\bar{Y}^2,$$

ahol kihasználtuk, hogy $\sum Y_i = N\bar{Y}$. A (4) formula felírásakor emellett még azt is figyelembe vettük, hogy $\frac{\sum Y_i^2}{N} = \bar{Y}_q^2$.

A *súlyozott* esetre vonatkozó $\sum f_i d_i^2 = \sum f_i Y_i^2 - N\bar{Y}^2$ összefüggés ehhez teljesen hasonló módon igazolható. A súlyozott esetben a (4) formulába természetesen mindkét átlagnak a *súlyozott* változata helyettesítendő.

4. Kiegészítő megjegyzés az aszimmetria P mutatójához

A 2.16. táblázat alatt megfogalmaztuk azt az empirikus összefüggést, hogy a mérsékelt aszimmetrikus eloszlások esetében a medián rendszerint – de nem mindig! – úgy harmadolja a módusz és az átlag közötti távolságot, hogy az átlaghoz esik közelebb. Ez egyben azt jelenti, hogy

$$\bar{Y} - Mo \approx 3 \cdot (\bar{Y} - Me).$$

Ezért az aszimmetria Pearson-féle mérőszáma valójában az $\bar{Y} - Mo$ különbséget veszi tekintetbe, de a fenti empirikus összefüggés alapján becsülve. Ennek az a fő előnye, hogy osztályközös gyakorisági sor rendelkezésre állása esetén a medián közelítő értéke rendszerint jóval megbízhatóbb mint a módusz közelítő értéke.