

Megjelent: Magyar Földrajzi Konferencia tudományos közleményei (CD), Szeged, 2001

A területi lehatárolások statisztikai következményei

A területi lehatárolások statisztikai következményeinek megközelítése több oldalról és módszerrel is lehetséges. Ebben a tanulmányban a kérdéskört elsősorban szakirodalmi példákon keresztül tárgyalom, ezen kívül pedig saját számításokkal illusztrálva mutatom be.

A lehatárolási probléma jellege

A területi statisztikai elemzések előfeltétele valamilyen térfelosztás, más szóval valamilyen módon lehatárolt térrészek létezése. A lehatárolási probléma jelentkezésének kétféle fokozatát célszerű megkülönböztetni (1. táblázat). Az első fokozatban a vizsgálatba vont területegységet különítjük el a vizsgálat során figyelmen kívül hagyott területektől. Ettől a mozzanattól csak a világ egészének vizsgálata során tekinthetünk el. A második fokozatra akkor van szükség, amennyiben a vizsgált térrészen belüli területi különbségekre is kíváncsiak vagyunk. Ekkor belső határvonalakat is ki kell jelölnünk.

1. táblázat A lehatárolási probléma két fokozata

A probléma fokozata	A probléma lényege	Jellegzetes határvonalak
1. fokozat	Vizsgált térrész elhatárolása	Országhatár
2. fokozat	Vizsgált térrészen belüli határvonalak megállapítása	Országhatár, közigazgatási egységek határa

Forrás: saját szerkesztés

Az első fokozatú problémának létezését itt csak inkább jelezzük, mintsem részletesebben foglalkozunk vele. A határvonalak meghúzása során társadalmi adatok elemzése esetén többnyire a már meglévő közigazgatási határokhoz igazodnak. Ezek közül az országhatárok és a településhatárok figyelembe vétele tekinthető a viszonylag legkisebb nehézséget jelentőnek. Amikor a területegységet önmagában, belső területi tagozódásától eltekintve vizsgáljuk, akkor a lehatárolási problémának csak az első fokozata jelentkezik, mégpedig úgy, hogy mennyire tekinthető indokoltnak az adott területegység vizsgálata. Mindig lehet állítani, hogy az illető területegység valamilyen szempontból kiemelt fontosságú, ekkor legfeljebb ezen indokokat lehet kritika alá vetni, nem magát a lehatárolást. Amikor valamilyen hatalmi szervezet területi hatóköre által meghatározott a vizsgálati egység, akkor ritkán érezzük problémásnak a határvonalak kérdését. Például országok esetén szinte mindig magától értetődőnek tartjuk, hogy releváns vizsgálati egységről van szó.

Léteznek azonban olyan helyzetek mind a társadalom-, mind a természetföldrajz területén, amikor az alaplehatárolás kérdése maga is élesen felvetődhet, például a

- határmenti térségek
- városok, objektumok vonzáskörzete
- vonalas infrastruktúrák térsége
- tájegységek

elemzése során nincsen egyértelmű módszer ezeknek a térségeknek az elhatárolására.

A továbbiakban az első fokozatú lehatárolási problémával nem foglalkozunk, feltételezzük, hogy létezik egy vizsgálati szempontból releváns térrész. A második fokozatú probléma statisztikai elemzésekre gyakorolt hatását nézzük meg behatóbban.

A területileg aggregált adatok elemzése az egyéni adatok elemzéséhez képest merőben új problémákat vet fel. A területi statisztikai adatok a területi ismérv alapján csoportosított adatok. A sajátos területi statisztikai problémák az adatok aggregált jellegével és a csoportosító ismérv, a területi dimenzió sajátosságaival függenek össze. Maga a lehatárolási probléma két részproblémára osztható, az aggregációs problémára vagy más néven a skálázás kérdésére és a módosítható egység problémájára.

Az aggregációs problémát az ökológiai korreláció példáján keresztül tárgyalom, mivel leggyakrabban ezen a néven foglalkoztak vele eredetileg elsősorban szociológiai indíttatású kutatások. Ezek legfőbb tanulsága, hogy a területileg csoportosított adatokból leszűrt következtetések csupán a területi egységekre vonatkoznak, az egyedekre nem. A módosítható egység problémája az ökológiai tévkövetkeztetésnél általánosabbnak, a területi statisztika legalapvetőbb problémájának tekinthető. Az elnevezés arra utal, hogy a területi felosztásokban nem beszélhetünk olyan alapegységekről, amelyek nem oszthatóak tovább, a lehatárolás változtatásával, ha nem is tetszés szerint, de azért manipulálhatóak az adatok, ráadásul előre nem kiszámítható mértékben. A két típusú probléma természetesen nem mindig különül el élesen az egyes tanulmányokban, a használt terminológia sem egyezik meg teljesen minden esetben, például a módosítható egység problémájához lehetne sorolni a területegységek összevonásának, vagyis aggregálásának kérdését is. Mégis célszerűbbnek tartom külön tárgyalásukat, egymással való szoros kapcsolatuk hangsúlyozása mellett. Olyan tanulmányokon keresztül mutatom be a problémákat, amelyek teljes egészében a kérdéssel foglalkoznak, de az egyes szerzőknek csak a leghangsúlyosabb és legfontosabb eredményeit és megállapításait ismertetem. Így az ismétléseket vagy a hasonló gondolatok bemutatását nem lehetett mindig elkerülni.

Az ökológiai korreláció problémája

Először a népszámlálási adatokkal kapcsolatos elemzések során mutatott rá Neprash az 1930-as években az Egyesült Államokban arra, hogy a területileg csoportosított jellemzők közötti korrelációk nem a változók közötti kapcsolat erősségét mérik, hanem csupán a földrajzi eloszlásuk közötti hasonlóságot vagy különbözőséget mutatják ki. További megszorításként a korrelációs együtthatókat csak akkor értelmezhetjük helyesen, hogyha a területi egységek egyenlő méretűek, vagyis egyenlő a területi kiterjedésük, amennyiben a területi eloszlásuk képezi a vizsgálat tárgyát, illetve egyenlő népességszámúak, amennyiben az emberek közötti gyakoriság a kérdéses. A körzetek méretének a csökkenésével növekszik a korrelációs együtthatók magyarázó ereje, egészen addig a szintig, ami alatt már nem csökkenthető a körzetek mérete. Külön problémát jelent az a kérdés, hogy a nagy méretű heterogén területeken számított korrelációs együtthatók mekkora jelentőséggel bírnak. A heterogén területek egy részén ugyanis jelentős korrelációt mérhetünk, más részén ugyanakkor pedig teljes korrelálatlanságot vagy ellentétes előjelű korrelációt lehet kimutatni. (Neprash, 1934)

A területegységek összevonásának hatását Gehlke és Biehl a fiatalok által elkövetett bűntettek száma és az átlagos havi jövedelem között mért korreláción keresztül illusztrálta. Cleveland 252 népszámlálási körzete alapján számított korrelációs együttható mértéke $-0,502$. Az egymással határos körzetek összevonása révén a korreláció mértéke fokozatosan erősödik, 25 körzet esetén $-0,763$ -ra. Ha a körzeteket véletlenszerűen vonják össze, függetlenül a földrajzi fekvésüktől, az együttható $-0,544$ lesz. A körzetek méretének változásával változik a korreláció mértéke, kisebb körzetekhez kisebb korrelációs értékek tartoznak. Ezek az eredmények kérdésessé teszik, hogy ezeknek a korrelációknak bármiféle szerepük is lehetne az oksági magyarázatok során, és hogy a jellemzők eredeti tulajdonosaira, az egyénekre és családokra is igazak legyenek. (Gehlke–Biehl, 1934)

Ezen felvetéseket követően a téma kutatásában az igazi áttörést Robinson alapvető jelentőségű 1950-ben megjelent cikke jelentette. Magát az ökológiai korreláció fogalmat is itt vezette be Robinson, akinek definíciója szerint az ökológiai korreláció tárgyat egyének csoportjai képezik. A feketék aránya az egyes USA tagállamokban és az írástudatlanság aránya közötti korreláció például ökológiai korreláció, szemben például az írástudatlanság és a bőrszín közötti egyéni korrelációval. Az egyéni korreláció tárgyai oszthatatlan egységek, mint az egyének, lakások, családok. Robinson rámutatott, hogy számos tanulmányban használnak teljesen tévesen ökológiai korrelációt az egyéni viselkedés leírására. Ökológiai korrelációt ekkor nem azért alkalmaznak, mert a csoportok viselkedése iránt érdeklődnek, hanem mert az egyéni korrelációk számítására adatok hiányában nincsen lehetőség.

Az ökológiai korreláció és az egyéni korreláció is az egyes területeken belüli egyéni korrelációktól függ, de eltérő módon. Az egyéni korrelációk az egyedek jellemzőinek területeken belüli gyakoriságaitól függenek, az ökológiai korrelációk viszont a jellemzők területen belüli gyakoriságait nem veszik figyelembe, csak a területek vizsgált jellemzőinek átlagait. Ugyanannak az átlagos értéknek azonban egymástól merőben eltérő egyedi értékek felelhetnek meg. Ezért egy ökológiai korrelációnak számos különböző egyéni korreláció felelhet meg, és viszont, az egyéni korrelációk is számos különböző ökológiai korrelációt eredményezhetnek az eltérő csoportosításoknak köszönhetően. Robinson rámutat arra is, hogy az ökológiai korrelációt lehet számítani súlyozott és súlyozatlan adatokkal is, a kettő közül a súlyozott számítás tekinthető korrektebbnek.

Robinson két példán keresztül illusztrálja a problémát. A feketék és az írástudatlanok aránya közötti ökológiai korreláció mértéke 1930-ban az USA tagállamok szintjén 0,773, a népszámlálás kilenc ország részének szintjén 0,946, az egyéni korreláció pedig 0,203, ami töredéke az ökológiai korrelációknak. De ebben az esetben legalább az előjel változatlan maradt, nem úgy, mint a külföldön született amerikaiak és az írástudatlanság közötti korreláció esetében, amikor az ökológiai korreláció -0,526 és -0,619 (állami és ország rész szinten), az egyéni korreláció viszont 0,118. Vagyis, mint az az adatokra vonatkozó keresztábrából is látszik, a külföldön született amerikaiak között magasabb arányban találunk írástudatlanokat, mint az USA-ban születetteknél, az ökológiai korreláció viszont pont ennek az ellenkezőjét sugallná.

Robinson levezeti, hogy az egyéni és az ökológiai korreláció akkor lesz egyenlő, ha az átlagos területeken belüli egyéni korrelációk nem kisebbek a teljes individuális korrelációnál. Ha a kisebb terület egységeket összevonjuk, az ökológiai korreláció értéke nagyobb lesz.

Összefoglalásként Robinson megállapítja, hogy az ökológiai korreláció nem alkalmas az egyéni korrelációk helyettesítésére. Bár elméletileg egybeeshet a két érték, ennek feltételei gyakorlatilag nem szoktak előfordulni. Robinson célja az volt, hogy az egyéni viselkedés vizsgálata kapcsán a jelentés nélküli ökológiai korrelációk használatát kerüljék el, és használjanak egyéni korrelációkat. (Robinson, 1950)

A Robinsont követően a témával foglalkozók egyetértenek abban, hogy az ökológiai korrelációk nem helyettesíthetik az egyéni korrelációkat. Robinson elemzését kiterjesztve megvizsgálták a dichotóm, mennyiségi és minőségi ismérvek közötti ökológiai kapcsolatok vizsgálatának sajátosságait. Goodman két esetet különböztet meg, amelyek során ökológiai korrelációkat szoktak alkalmazni. Az egyik esetben két jellemző földrajzi eloszlásának megegyezőségét, illetve különbözőségét szeretnék egy mutatószámmal kifejezni, például ha valakit a feketék és az írástudatlanok földrajzi eloszlásának hasonlósága érdekel az Egyesült Államokban az államok szintjén vagy más területi felosztás szerint. Ekkor indokolt használatuk. A másik esetben az ökológiai korrelációk segítségével következtetni szeretnének az egyéni viselkedésre, illetve korrelációkra. Ezek a vizsgálatok azonban általában hibásak, csupán rendkívül speciális körülmények esetén lehetne az ökológiai korrelációt felhasználni az egyéni viselkedés leírására. (Goodman, 1953)

Duncan és Davis tanulmányában egyetért Robinson megállapításával, de rámutatnak arra, hogy a területileg csoportosított adatok vizsgálata során az ökológiai korrelációk számítása nem a legjobb elemzési módszer, pontosan azon tulajdonság miatt, hogy értékük széles sávban mozoghat. Három népszámlálási adaton alapuló példán keresztül mutatják be, hogy a népszámlálási körzetenként ismert megoszlások alapján az egyéni korrelációk elméletileg milyen széles tartományban helyezkedhetnek el. Részletesebb térfelosztás esetén a sáv szélessége csökken. (Duncan–Davis, 1953)

Sawicki a témáról adott áttekintésében Blalockot követően hangsúlyozza, hogy Robinson problémája egy általánosabb kérdés, az aggregációs probléma speciális részét alkotja, mégpedig azt az esetet, amikor az aggregálás kritériumaként a területi folytonosság jelenik meg. A földrajzi csoportosítás hasonlít a véletlenszerű csoportosításhoz, amikor a csoportosítási eljárás korrelációs együtthatóra gyakorolt hatását szinte lehetetlen előre jelezni. Az aggregált adatok vizsgálata különböző szinteken történhet, a konkrét szituációtól és a felmérés céljától függetlenül lehetetlen megállapítani a legjobb, optimális szintet. Mindegyik szintnek megvan a maga létjogosultsága. A szomszédsági, városrészi vagy városok közötti különbségeket vizsgáló tanulmányok más és más oksági magyarázatát tárhatják fel ugyanannak a jelenségnek. Az aggregált számítások eredményei semmilyen más aggregáltsági szinten nem érvényesek. Például ha a kisebb jövedelmű körzetekben nagyobb a republikánus szavazók aránya, az nem jelenti azt, hogy a kisebb jövedelmű emberek között nagyobb a republikánus szavazók száma.

Egy hipotézis vizsgálata különböző területi lehatárolásokkal történhet, amelyek eltérő eredményekre vezethetnek. Ez felveti azt a tyúk-tojás problémához hasonló kérdést, hogy mit kell vizsgálni először, a területi lehatárolásokat, amelyeket adatok hiányában nem lehet elvégezni, vagy az adatokat, amelyek viszont területi lehatárolás nélkül még nem léteznek. Sawicki Syracuse városára vonatkozó társadalmi adatok vizsgálata során az eltérő aggregáltságú szinteken részben eltérő kapcsolatokat mutatott ki, például a népszámlálási körzetek szintjén a nagyobb tulajdon nagyobb társadalmi interakcióval párosult, míg az alacsonyabb szinteken, részletesebb térfelosztás mellett ilyen kapcsolat már nem volt kimutatható.

A csoportosított adatok vizsgálatát nem pusztán az a kényszer szüli, hogy adatok hiányában gyakran nincsen lehetőség egyéni adatok vizsgálatára, hanem az is, hogy bármiféle beavatkozás számára a csoportok jellemzői könnyebben elérhetőek, mint az egyéneké. Egyéni jellemzők gyakran változtathatatlan, objektív adottságok, például az egyén neme, kora, bőrszíne. A csoportok jellemzői ezzel szemben alakíthatóak. Az ökológiai korrelációk nem helyettesítik az egyénieket, hanem más célt szolgálnak. A számítások nem jelölik ki a cselekvés konkrét irányát, mivel ugyanaz az aggregált mutató az átlagok ismert tulajdonságai miatt számos módszerrel megváltoztatható. (Sawicki, 1973)

Openshaw számos új vizsgálati szempont felvetése mellett újabb empirikus vizsgálatokkal támasztotta alá az ökológiai korreláció fontosságát. Az Egyesült Királyságban és a legtöbb további országban is a népszámlálási adatok csupán területileg aggregált formában férhetőek hozzá, lényegében önkényes földrajzi körzetek szintjén. A népszámlálási körzeteknek viszont nincsen természetes vagy lényeges földrajzi tartalma. Az aggregálás adatokra gyakorolt hatásáról viszonylag kevés tanulmány született és az eredmények értelmezése során is meglepően kevés figyelmet fordítanak rá. A probléma viszont nagyon lényeges, és általánosságban is felveti a népszámlálási adatok használhatóságának kérdését.

Az ökológiai korreláció gyakorlati jelentőségét két elérhető adatbázison keresztül illusztrálja Openshaw. Az egyik a sunderlandi háztartások 10%-áról készült felmérés, a másik Firenze háztartásainak adatai. Az egyik esetben 53, a másikban 40 különböző változó közötti egyéni és ökológiai korrelációt számított ki és hasonlított össze Openshaw. (2. táblázat) Az ökológiai korrelációkat Sunderlandban egy, valamint fél kilométeres négyzet alakú körzetekkel,

Firenzében pedig a népszámlálási körzetek alapján csoportosítva számította. A terület egységek méretének csökkenésével azok homogénebbé is válnak, így az egyéni és ökológiai korrelációk közötti különbségek csökkentek. Az aggregáció hatása nem előrejelezhető és matematikai eszközökkel nem korrigálható.

2. táblázat Egyéni és ökológiai korrelációk keresztábrája (a soronkénti összes korreláció százalékában)

Egyéni korrelációk: -tól/-ig	Ökológiai korrelációk: -tól/-ig										Összesen
	-1.0/ -0.8	-0.8/ -0.6	-0.6/ -0.4	-0.4/ -0.2	-0.2/ 0.0	0.0/ 0.2	0.2/ 0.4	0.4/ 0.6	0.6/ 0.8	0.8/ 1.0	
Sunderland 1 kilométeres négyzetek (53 változó)											
-1.0/-0.8	100										1
-0.8/-0.6	50	50									4
-0.6/-0.4	12	44	32	12							25
-0.4/-0.2		9	36	34	15	4	1				180
-0.2/0.0			4	32	39	18	5	1			997
0.0/0.2			1	2	14	29	32	20	3		188
0.2/0.4						14	32	39	14		28
0.4/0.6							17	50	17	17	6
0.6/0.8								50	50		2
Összesen	6	32	117	387	444	248	117	66	13	1	
Firenze népszámlálási körzetek (40 változó)											
-1.0/-0.8	100										1
-0.8/-0.6		0									0
-0.6/-0.4			100								2
-0.4/-0.2	2	19	31	24	17	6					83
-0.2/0.0		1	7	21	32	23	14	2			603
0.0/0.2			1	6	10	28	28	22	3		78
0.2/0.4							18	27	55		11
0.4/0.6								100			1
0.6/0.8									100		1
0.8/1.0											0
Összesen	3	21	72	154	214	167	106	33	10	0	

Forrás: Openshaw, 1984

A csoportosítás hatása természetesen nemcsak a korrelációval kapcsolatban lép fel, hanem bármilyen más statisztikai módszer kapcsán is. A csoportosított népszámlálási adatok nem tesznek különbséget például a szegény emberek és szegény körzetek között. Ezt jól illusztrálja a firenzei adatokkal elvégzett kétféle klaszteranalízis összehasonlítása is. Az egyik csoportképzést a népszámlálási körzetek, a másikat egyéni adatok szerint végezték. A kétféle összehasonlítása alapján megállapítható, hogy az egyes elemek csoportjai között nincsen a minimálisan szükségesnél nagyobb átfedés. Ez nem meglepő, mivel a klasszifikált egységek – háztartások, illetve önkényesen meghúzott népszámlálási körzetek – a két esetben különbözőek. A klasszifikáció is más jellemzők szerint történt, az egyéni adatok szerint például az életkornak csoportképző szerepe volt, a területileg összesített adatokban pedig nem. A területi adatokkal végzett elemzés a hangsúlyt az átlagos adatokra helyezi, amelyek az adatok nagy szórása esetén nem tekinthetők tipikusnak, az egyedeket jól jellemzőknek.

Openshaw szerint az ökológiai tévkövetkeztetés lehetősége függ a vizsgálat módszerétől és az eredmények interpretálásának módjától. Nincs ismert módszer, amely az aggregálás hatását képes lenne mérni. A 2. táblázat extrém értékei például rendkívül félrevezetőek lehetnek, azok

alaján téves következtetésekre lehet jutni. Ha individuális korreláció számítására az adatok hozzáférhetőségének hiányában nincs mód, akkor a torzító hatás nagyságáról nincs ismeretünk. (Openshaw, 1984)

A módosítható egység problémája

Egyes területi adatok jellegüknél fogva nem egyedi adatok aggregációjaként állnak elő, hanem területi kiterjedésre vonatkozó fajlagos vagy átlagos értékeként. Ilyen jellemzők például a népsűrűség, a növények termésátlaga, a csapadékmennyiség, az árszínvonal, az útvonalsűrűség. Az ilyen típusú adatoknál nincs értelme egyedi adatokról beszélni, mivel természetüknél fogva valamilyen módon meghatározott területegységekre vonatkozó átlagokról van szó, az átlagok pedig mindig aggregátumokból származnak. A földrajzi területegységeknél pedig nincsen természetes egységük, határvonaluk módosítható.

A módosítható egység problémáját először Yule és Kendall fogalmazta meg és illusztrálta is a burgonya és búza termésátlaga közötti korrelációval. A 48 angol megye adatai alapján a két termésátlag közötti korreláció 0,2189. A 48 megyét 24 megyévé összevonva az együttható értéke 0,2963-ra növekszik, 12 tag esetén 0,5757, hat területegységnél 0,7649, háromnál pedig 0,9902 lesz a korreláció mértéke. Területileg és időbelileg folytonos terjedelmű változók esetében mindig fellép a módosítható egység problémája. A korrelációk együtthatók a változók közötti kapcsolatot a választott egységekre vonatkoztatva tükrözik. „Érvényességük nem független ezektől az egységektől, hanem viszonylagos. Nem csupán a vizsgált tulajdonságok változásait mérik, hanem a rendszeres mérés céljából rákényszerített hálózati felosztás egységének tulajdonságait is.” (Yule–Kendall, 1964, 321. oldal) Ugyanakkor ezek a korrelációk természetesen nem jelentés nélküliek, csak egyszerűen nem általános érvényűek. (Yule–Kendall, 1964)

Yule és Kendall tökéletesen bemutatta a probléma jellegét, hasonlóan ahhoz, mint ahogyan Robinson bemutatta az ökológiai tévkövetkeztetés lényegét. A későbbi kutatások itt is a további finomításokat, illusztrációkat szolgálták. A téma legtermékenyebb kutatója Openshaw, aki szerint a skálázási probléma azért lép fel, mert a területi adatokkal kapcsolatban mindig bizonytalan az a kérdés, hogy a vizsgált területet hány zónára osszuk fel, az aggregációs probléma pedig azért, mert ezen adatok aggregálása számos módon megvalósítható. A területegységek határvonala általában véletlenszerűnek mondható, amelyek többnyire a hagyományokon és az adatok elérhetőségén alapulnak. Ezt kitűnően mutatja Openshaw következő példája. Egy South Tyneside-i körzet házaira vonatkozó adatbázis 1219 százszor száz méteres négyzetek formájában áll rendelkezésre. Ezek a négyzetek többféle módon aggregálva lettek 200, 300, 400, 500, 600, 700, 800, 900, 1000 méter oldalú négyzetekké. Az 500 méteres négyzeteket például 25 különböző módon lehetett létrehozni. A különféle négyzethálókra kiszámolta a korai viktoriánus és a közép viktoriánus házak közötti korrelációt, ezeknek az átlagát és szórását. (3. táblázat)

3. táblázat Az aggregáció hatása a korai viktoriánus és a közép viktoriánus házak közötti korrelációra

Négyzet nagysága	Korreláció átlaga	Szórás
200 méter	0.315	0.112
300 méter	0.438	0.069
400 méter	0.478	0.113
500 méter	0.494	0.165
600 méter	0.523	0.163
700 méter	0.572	0.182
800 méter	0.582	0.184
900 méter	0.605	0.197
1000 méter	0.626	0.201

Forrás: Openshaw, 1977

A területi alapegységek a legkisebb területi egységek, amelyekre adatokkal rendelkezünk, ezeket az aggregálás során alapadottságnak kell elfogadnunk. A fenti példában a 100*100 méteres négyzetek az alapegységek, amelyek ugyanúgy önkényesek, mint ahogyan bármely más területű síkidom is az lenne. Lehetne az alapegység 14*14 méteres négyzet vagy 22 méteres oldalú szabályos háromszög. Openshaw többféle eljárást bemutat, amellyel a területi alapegységek zónákba aggregálása bizonyos, általunk megszabott feltételek szerint történik. Az aggregálási kritérium lehet minimális vagy maximális szórása bizonyos jellemzőnek, korreláció minimális vagy maximális értéke és bármi más. Így meg lehet állapítani az eredmények lehatárolási módra vonatkozó érzékenységet. Az egyik példában a két világháború között és a második világháború után épült házak közötti korreláció maximumát számolta ki különféle zónaszámok esetén hierarchikus heurisztikus eljárással. (4. táblázat) A következő példában egy másik eljárást alkalmazva szintén a maximális pozitív korreláció elérése volt a cél. (5. táblázat) A 400 méteres oldalú négyzetekkel kiindulva a minimális korreláció $-0,949$, ami összehasonlítva az 5. táblázat $0,999$ értékével jelzi a zónázási (aggregálási) módszer potenciális fontosságát. Egy további számításban a cél a lineáris regressziós modell legjobb illeszkedése volt, ami az eredeti 25-30%-os értékek után a megfelelő aggregálási módnak köszönhetően 99,9%-osra növekedett. (Openshaw, 1977)

4. táblázat Maximális korreláció a két világháború között és a második világháború után épült házak között hierarchikus heurisztikus eljárással számolva

Zónák száma	Korreláció
1000	0.539
900	0.647
800	0.766
700	0.863
600	0.902
500	0.904
377	0.905
195	0.891
123	0.865
93	0.868
64	0.821
52	0.820
40	0.803
30	0.772

Forrás: Openshaw, 1977

5.táblázat Maximális pozitív korreláció a két világháború között és a második világháború után épült házak között automatikus zónázással számolva

Négyzetek eredeti nagysága	Zónák eredeti száma	Zónák végső száma	Eredeti korreláció	Végső korreláció
200 méter	377	10	-0.327	1.000
300 méter	195	49	-0.273	1.000
400 méter	123	29	-0.233	0.999
500 méter	93	28	-0.143	0.999
600 méter	64	24	-0.140	0.999
700 méter	52	24	+0.006	0.999
800 méter	40	28	-0.124	0.999
900 méter	40	21	-0.108	0.999

Forrás: Openshaw, 1977

Openshaw további munkákban is foglalkozik a zónázás hatásával, mivel szerinte az eddigi vizsgálatok többsége megmarad verbális szinten, ami csak a jelenség elvi fontosságát tudja feltárni, gyakorlati jelentőségét viszont nem. Minden vizsgálata szerint a zónázási módszer fontos befolyást gyakorol az adatokra és a velük végzett műveletekre. Taylornal írt közös tanulmányában kitűnő áttekintést nyújt a kérdés elméleti vonatkozásairól. Mint itt írják, három okát lehet megkülönböztetni annak, hogy egyes kutatók a módosítható egység problémáját nem veszik komolyan. Először hivatkozhatnak a megoldhatatlanságára, másodsor trivialitására és lényegteleniségére, harmadszor pedig arra, hogy létezése kétséget támaszt szinte az összes területi adattal kapcsolatos mennyiségi elemzési módszer alkalmazhatóságával kapcsolatban. Bizonyítottnak lehet tekinteni, hogy az első két hivatkozás hibás, helytelen, míg a harmadik nagyrészt igaz. A mennyiségi elemzésekről való lemondás azonban nyilvánvalóan nem lenne hasznos megoldás, hiszen ezeket semmi mással nem tudnánk pótolni. Ehelyett arra van szükség, hogy a területi kérdésekkel foglalkozó kutatók tisztában legyenek a problémával és az alkalmazott eszközök korlátaival. Azon területi kutatások komoly hiányosságokkal bírnak, amelyek nem veszik figyelembe a módosítható egység problémáját. Ezek a munkák lehetnek statisztikai vagy matematikai értelemben magas színvonalúak és jók, de mint területi kutatások mégis rosszak.

Minden területi kutatásnak meg kell válaszolnia azt az explicit formában egyébként gyakran nem felvetődő kérdést, hogy milyen területi egységekkel folyjék. A területi felosztással válik elemezhetővé a területi szerkezet. Amennyiben az elemzés érzéketlen lenne a területi felosztás módjára, akkor nem lenne a kérdés problémás. A területi egységek egy része a priori módon adottnak tekinthető. Ilyen egységek a helyi önkormányzatok, a területi középszintek, mint megyék, tartományok stb., a népszámlálási körzetek, a postai irányítószámok, a választókerületek. Ezeknek nagy része a földrajzi, területi kutatások számára irrelevánsak. Mégis használják őket ilyen célokra, amelynek a kényelmesség és az adatok elérhetősége mellett az a nézet is oka, miszerint a területi egységek megválasztása lényegtelen szerepet játszik a kutatásokban. A területi egységek fontosságának észlelése esetén a kutatók egy része új, saját lehatárolás megalkotása mellett dönt. Ennek a legegyszerűbb módja a geometriai alakzatok (négyzetek, hatszögek) használata. A bonyolultabb módszerek valamilyen algoritmus segítségével, a tapasztalati adatok felhasználásával képeznek régiókat.

A módosítható egység problémájának kezelésére egyes kutatók eddig is többféle módszert javasoltak. Az egyik eljárás önkényes lehatárolási kritériumok szerint végzi el a térfelosztást, például azonos lakosság számú egységeket hoznak létre, vagy a területi áramlás vizsgálatánál minimalizálják a terület egységeken belüli áramlást. Az ilyen módszereket félmegoldásnak

sem lehet tekinteni, mert ugyanúgy eltekintenek az önkényes lehatárolás eredményekre gyakorolt hatásától, mint az a priori terület egységek használata esetén.

Néhány kutató, mint például Cliff vagy Williams, analógiát vonnak a módosítható egység problémája és a mintavételből származó mintavételi hiba kérdése között. Ezen nézet szerint a skálázás (a terület egységek számának megállapítása) a mintanagyság kérdésével állítható párhuzamba, a módosítható egység problémája pedig a mintavételi hibával. Ez az analógia érvényessége esetén kitűnő megoldás lenne, hiszen statisztikailag szilárd alapot teremtene a különféle térfelosztási módszerekből származó eltérő eredmények kezelésére. Ez a hatás standard formulák segítségével lenne mérhető, hasonlóan a mintavételi hiba nagyságának becsléséhez. A két kérdés között azonban alapvető a különbség, mert a mintavételi hibát többnyire meg lehet becsülni, ezzel szemben az adatok zónázási módból fakadó változékonysága nagyrészt ismeretlen mértékű.

Egy további nézet szerint a módosítható egység problémája csak azért létezik, mert bizonytalanság uralkodik abban, hogy mik legyenek a területi kutatások területi egységei. Ha ezt a bizonytalanságot megszüntetjük, eltűnik a probléma. Csupán arra van szükség, hogy bár szubjektív és önkényes, de mindenki által elfogadott lehatárolási keretek között folyjanak a kutatások. Az ilyen egységek azonban távolról sem felelhetnének meg a sokféle elemzési célnak, csupán az adatok egyszerű közlésére alkalmasak.

Egy másik megközelítés úgy kezeli a kérdést, hogy a területi alapegységek zónába egyesítése során a kívánt eredményekhez igazítja a területi lehatárolást. Például lehet a kritérium két jellemző közötti maximális korreláció vagy adott zónaszám melletti maximális szórás vagy bármilyen más konkrétan megállapított eredmény. Ennek a módszernek a segítségével lehet tesztelni különféle hipotéziseket és meg lehet állapítani az eltérő zónázási módszerek szélsőértékeit. (Openshaw–Taylor, 1981)

Bach empirikus vizsgálata arra vonatkozott, hogy az intézmények telephelyválasztási döntéseinél milyen szerepet játszik az aggregációs szint és a használt távolságfogalom. Az aggregációs szintnél meg kell különböztetni a kínálat szintjét, vagyis azt, hogy a potenciális telephelyek kijelölésénél milyen részletes legyen a térfelosztás, és a kereslet szintjét, azaz annak eldöntését, hogy az intézmények iránti kereslet megállapításánál milyen térfelosztást alkalmazunk. Bach Dortmundban és Kleve-Emmerichben vizsgálta meg azt, hogy különféle számú szolgáltatóintézmény optimális telephelyválasztása változik-e, amennyiben négy eltérő aggregációs szinten végzi el a számításokat. A lehatárolás változtatásának hatását Bach nem vizsgálta, de eredményei így is érdekesek. Az egyes aggregációs szinteken végzett számításokban a potenciális telephelyek helyzete nem változott, csak a kereslet helyzete tért el a terület egységek összevonása miatt. Az eredmények szerint az aggregációs szintek változásával az optimális telephelyek csak kismértékű hasonlóságot mutatnak. Amennyiben hét intézmény telephelyének kijelölése a cél, akkor a szintek egybevetése alapján csak kettő, illetve a másik esetben három makrorégiót lehet telephelynek elkülöníteni, a többi optimális helyzetűnek számított telephely egymástól túlságosan szétszórva található. Ebből az szűrhető le, hogy az aggregációs szint komoly befolyást gyakorol az optimális telephelyek kijelölésére. Bach szerint az eredmények különbözősége akkor nem jelenthet problémát, hogyha az egyes intézményeknek megfelelő aggregációs szinten végzik el a telephelyek kijelölését. Az óvodák számára például a kereslet a szomszédság szintjén, az uszodák esetében kisvárosi vagy városrészi szinten, a kórházaknál pedig megyéni területek szintjén jelentkezik. (Bach, 1981)

Fotheringham, Densham és Curtis a zónázási rendszer telephelyválasztási modellek eredményére gyakorolt hatását vizsgálta. Elemzésük Bachénál részletesebb, mert nem csupán az aggregációs szintek, hanem a határvonalak változtatásának jelentőségét is megnézték. A telephelyválasztási modelleknek Hillsman és Rhoda alapján három speciális hibaforrását lehet elkülöníteni, amelyek az aggregált adatok használata miatt lépnek fel. Az „A” típusú hiba abból a feltételezésből ered, hogy a területi kereslet terület egységenként csupán egyetlen

pontban jelentkeznek a pontok halmaza helyett. A „B” típusú hiba oka a területegységen belüli távolságok figyelmen kívül hagyása, nulla értékkel számítása. A „C” típusú hibát az okozza, hogy az egyes területek kereslete csupán egy szolgáltatóhely felé irányulhat, pedig a területegységek egyes részei más-más szolgáltatóhelyhez lehetnek közelebb. Ezeknek a problémáknak az előrejelzése, számszerűsítése alig lehetséges. Végül, a három egymással összefüggő speciális hibaforráson kívül számolni kell a területegységek lehatárolási módjának és aggregálásának, vagyis a módosítható egység problémájával is, amit zóna definiálási problémának neveznek a szerzők. Ez azért különösen fontos a telephelyválasztási modelleknél, mert a modellek eredményeként kapott telephelyek optimálisnak vannak számon tartva, és mint ilyenek, komoly súllyal esnek latba a gyakorlati döntéshozatalnál is. Amennyiben a végeredmény egyszerűen befolyásolható a lehatárolás és/vagy az aggregáció változtatásával, akkor az optimális szó használatával megtévesztő volta miatt óvatossággal kellene bánni.

A szerzők gyakorlati számításai Buffalo nagyvárosi körzetének 871 népszámlálási körzetének alapadataiból indultak ki. A képzeletbeli cél tíz időségek központjának a létesítése volt, azzal a kritériummal, hogy a 65 éven felüli lakosság számára az összesített utazási távolság minimális legyen. A 871 körzet alapján kiszámolt központok képezték az alaperedményt, amihez a többi számítás hasonlították. Az alaperedményen kívül hat területi szinten (800, 400, 200, 100, 50, 25 zóna) egyenként 20 különböző lehatárolási móddal, vagyis összesen 120-szor futtatták le az algoritmust, így a szintek és a lehatárolás változtatásának hatását is vizsgálni tudták. Az aggregáció szintje és az átlagos távolság között negatív a kapcsolat, amit az a feltétel eredményez, hogy a területegységen belüli távolságoktól eltekintenek. A zónák nagyobbá válásával egyre nagyobb lesz annak a népességnek az aránya, amely a zónákon belül utazik a központokba. Amennyiben az eredeti 871 zónával számoljuk ki az átlagos távolságot, akkor bár mindegyik szinten bizonyos szórással, de egymáshoz közeli átlagos utazási távolságokat kapunk.

A 120 eredményt egyetlen térképen ábrázolva rendkívül meggyőző módon érvelhetünk az eredmények aggregációs szint és lehatárolási módtól való függése mellett. Optimális telephelyeket a Buffalo határához közeli sávot leszámítva az egész vizsgálat alá tartozó területen mindenhol szétszórva találunk. Ugyanakkor bizonyos kisebb körzetek jól láthatóan gyakrabban kerültek kiválasztásra, hét zónát lehet megjelölni, ahol sűrűbben fordulnak elő optimális pontok. Ez az eredmény mindenesetre bármelyik egyedi megoldás általános érvényűségét kétségbe vonja. Egyes megoldások a belvárost látnák el jobban, mások a város egyéb részei számára lennének kedvezőbbek. Az optimális szó tehát csak rendkívül korlátozott érvényű lehet, arra vonatkozik, hogy adott lehatárolás és aggregációs szint mellett mit tekinthetünk optimálisnak. (Fotheringham–Densham–Curtis, 1995)

A módosítható egység problémáját bármilyen elemzési eljárással kapcsolatban lehet vizsgálni. Fotheringham és Wang szerint a probléma egy és kétváltozós elemzésekre gyakorolt hatása viszonylag könnyen érthető pusztán elméleti megfontolások segítségével is, a többváltozós elemzéseknél azonban annyira komplexsége válik, hogy csak az empirikus munkák segíthetik elő bemutatását és megértését. A szerzők tudomása szerint a témában elméleti munka még nem is született és nem is nagyon várható. Mint azt 1990-ben írták, a számítógépekhez való könnyű hozzáférés miatt a többváltozós elemzések gyarapodására lehet számítani, ezért is szükséges a potenciális felhasználók figyelmét felhívni a problémára és óvatosságra inteni őket.

A szerzők két többváltozós regressziós modell eredményeire vonatkozóan vizsgálták meg, hogy a lehatárolási és aggregációs probléma mennyire jelentős. Az alapadatok Buffalo nagyvárosi körzetének 871 népszámlálási körzetére vonatkoztak. Az egyik modell függő változója az átlagos területi családi jövedelem, a magyarázó változók a lakástulajdonosok aránya, a fizikai munkások aránya, a feketék aránya, a 65 évesnél idősebbek aránya. A másik

modell ugyanezeket a változókat tartalmazta, de a függő változó a lakástulajdonosok arányának logaritmus volt. A regressziós egyenleteket az alapeseten kívül hat más szinten, 800, 400, 200, 100, 50, 25 zónaszámmal, egyenként 20 különböző lehatárolási móddal számították ki. Az aggregációs szint növekedésével két negatív paraméter – az idősök és a fizikai munkások aránya – értékei szisztematikusan tovább csökkennek. Ha például az idősök aránya 1%-al növekszik, akkor a családi jövedelem 871 egység alapján kiszámítva 3075 dollárral csökken, 25 egység esetén pedig átlagosan 26540 dollárral. A másik kettő paraméter a zónázási rendszer változtatására reagált érzékenyebben. A lakástulajdonosok arányának 1%-os változása 25 egység esetén 10430 dollár és 17910 dollár közötti értékkel változtat a családi jövedelmen. A determinációs együtthatók értékei a magasabb szinteken nagyobbak, 800 zóna esetén átlagosan 40%-os, 100 zóna esetén 85%-os az értékük, ezután már csak kisebb mértékben növekednek. Úgy látszik, az adatok megfelelő csoportosításával bármilyen kívánt megbízhatósági szint elérhető. A lehatárolás változtatásának hatását a szerzők úgy is megvizsgálták, hogy 150 különböző módon aggregálták 218 zónába az alapadatokat. Az eredmények szórása a fizikai munkások és a feketék esetében a legnagyobb, a fizikai munkásoknál a paraméter értéke 0 és 20000 dollár között található. Talán még érdekesebb, hogy az idősök aránya és a családi jövedelem között az esetek egy részében pozitív, más részében negatív volt a kapcsolat, és további esetekben pedig nem lehetett kapcsolatot kimutatni közöttük.

A szerzők szerint a többváltozós modellek rendkívül érzékenyek az aggregációs szintekre és zónázási módokra. A modellek készítőinek feltétlenül figyelemmel kell lenniük erre. Míg az egy- és kétváltozós módszereknél a módosítható egység problémájának hatása bizonyos mértékig előrejelezhető, a többváltozós elemzéseknél ugyanez lényegében előrejelezhetetlen. A probléma kezelése érdekében három megoldást vizsgálnak meg. Először is az eredményeket több aggregációs szinten és többféle lehatárolással célszerű kiszámítani, így lehet következtetni az egyes paraméterek stabilitásának mértékére. A másik megoldás az aggregált adatok helyett az egyedi adatok használata, amennyiben azok rendelkezésre állnak. Ez nagyon kevés esetben kivitelezhető, vagy adatvédelmi okok miatt, vagy mert egyszerűen nem létezik az adatnak alapegysége. A harmadik lehetőség az optimális zónázási rendszer kialakítása lenne, ez azonban a több változó miatt azért nem oldható meg egyértelműen, mert amelyik lehatárolás az egyik változó számára optimális, mert például teljesíti a maximális zónák közötti szórás kritériumát, az nem biztos, hogy a másik változó számára is az. A vizsgálatok tanulsága, hogy egyetlen lehatárolási mód eredményeinek felhasználása nem elégséges megalapozott következtetések levonására. (Fotheringham-Wong, 1991)

A területi lehatárolások hatása gyakorlati példák alapján

Négy gyakran használt statisztikai elemzési eszközzel végzett számításokkal mutatom be a lehatárolási probléma, azon belül elsősorban a területi szintek jelentőségét. Ezek az eszközök a maximális és minimális érték hányadosa, a súlyozott relatív szórás, a Hoover-index és a korrelációs számítás.

A maximális és minimális érték hányadosának rendkívüli előnye a könnyű számíthatósága és érthetősége. Könnyen belátható, hogy a területi szintek növekedésével a teljesen homogén területegységek összevonásának valószínűtlen esetétől eltekintve a területi különbségek csökkennek, mivel az átlagolódás következtében a maximális értékek csökkennek, a minimális értékek növekednek. Az egy lakosra jutó személyi jövedelemadó alap maximális és minimális értékeinek hányadosa a 6. táblázatban látható módon alakult.

6. táblázat Az egy lakosra jutó személyi jövedelemadó alap maximális és minimális értékeinek hányadosa

Területi szint	Területegységek száma	1991	1996
Település	3067	21,9	47,9
Kistérség	150	2,9	3,4
Megye	20	2,2	2,4
Régió	7	1,7	1,8
Főváros-vidék	2	1,7	1,7

Forrás: saját számítás a PM adatai alapján

A területegységek számának növekedésével növekednek vagy változatlanul maradnak, de biztosan nem csökkennek a területi különbségek, hogyha az addig meglévő határvonalak helyzete változatlan marad. A határvonalak változása esetén tendenciaszerű növekedésről beszélhetünk, ami időnként megtörhet a lehatárolás változásának hatására. Ezt lehet érzékeltetni egyre növekvő részletezettségű Dirichlet-poligonos térfelosztással. A 7. táblázat adatai olyan Dirichlet-poligonos lehatárolás mellett születtek, ahol a pontokat mindig települések középpontjai jelölték ki, a települések kiválasztása pedig nagyság szerinti sorrendben történt. Vagyis például a négy területegységes felosztás pontjai Budapest, Debrecen, Miskolc és Szeged középpontjaival egyeznek meg, az ötös felosztásnál ezen városok mellett Pécs is megjelenik poligonközpontként. A területi különbségek növekvő tendenciája egyértelműen leolvasható az adatokból. Érdekes azonban, hogy a telefonvonalak területi különbsége 60 poligon esetén nagyobb, mint 80 vagy 100 poligonnál. Ez a csökkenés azonban csak átmeneti. A másik érdekesség, hogy 25 poligonig az egy lakosra jutó személyi jövedelemadó alap különbségei nagyobbak a személygépkocsiénál, aztán 30, 50, és 60 poligonnál a személygépkocsiénál nagyobbak a területi különbségek, majd 80 és 100 poligonnál ismét a személyi jövedelemadó alapnál. Ez a jelenség arra hívja fel a figyelmet, hogy különböző lehatárolásoknál más és más lehet két jellemző területi különbségének egymáshoz való viszonya, egyes lehatárolások eredményeiből más lehatárolásokéra nem mindig lehet következtetni.

7. táblázat Egyes jellemzők maximális és minimális értékeinek hányadosa különböző térfelosztások esetén

Dirichlet poligonok száma	Egy lakosra jutó, 1996-ban		
	Személygépkocsi	Telefon	Szja alap
2	1,41	1,52	1,39
3	1,44	1,79	1,45
4	1,47	1,75	1,57
5	1,50	1,82	1,63
10	1,71	2,20	1,96
15	1,69	3,56	1,99
20	1,77	3,65	2,03
25	2,10	3,63	2,16
30	2,30	3,76	2,20
50	2,58	5,78	2,27
60	2,64	7,17	2,29
80	2,75	6,65	2,81
100	2,70	7,03	3,38

Forrás: saját számítás a KSH és a PM adatai alapján

A súlyozott relatív szórás szintén csökken a területi szintek növekedésével (8-9. táblázat). A megyék régiókká történő összevonása esetén például az azonos régiókba összevont megyék eltérései eltűnnek – belső szórásként kiszámíthatóak ezek az értékek is –, ami csökkenti a szórást. A főváros-vidék szintű szórás azért nagyobb, mint a regionális beosztással számított, mivel a megyei szintű szóráshoz képest Budapest és Pest megye összevonása erőteljesebben csökkenti a szórás mértékét, mint a Budapest kiugró értékéhez képest viszonylag kiegyensúlyozottabb összes megye összevonása és Budapest külön területegységként hagyása. A súlyozatlan szórás nem minden esetben csökken a területegységek összevonása esetén. A súlyozott szórás viszont a maximális-minimális értékek hányadosához képest egyenletesebben növekszik a területegységek számának növekedésével, mivel ez a mutató nem csak a szélsőértékeket veszi figyelembe.

8. táblázat Az egy lakosra jutó személyi jövedelemadó alap súlyozott relatív szórása, %

Területi szint	1991	1996
Település	32,7	35,7
Kistérség	28,0	30,5
Megye	24,3	26,4
Régió	20,2	22,3
Főváros-vidék	22,7	23,4

Forrás: saját számítás a PM adatai alapján

9. táblázat Egyes jellemzők súlyozott relatív szórása különböző térfelosztások esetén, %

Dirichlet poligonok száma	egy lakosra jutó, 1996-ban		
	személygépkocsi	telefon	szja alap
2	13,5	16,3	13,1
3	14,3	18,5	14,1
4	14,4	17,7	16,6
5	15,1	19,4	18,2
10	17,4	24,1	20,4
15	18,7	27,0	21,1
20	19,1	28,7	22,0
25	19,8	30,6	23,2
30	20,8	31,7	24,3
50	22,1	34,1	26,1
60	22,2	35,0	26,6
80	22,7	35,6	27,0
100	23,3	36,2	27,9

Forrás: saját számítás a KSH és a PM adatai alapján

A Hoover-index az előző kettő mutatóhoz hasonlóan viselkedik, nagyobb területi aggregáció esetén csökken a mértéke, kisebbek lesznek a területi különbségek az összevont területegységeken belüli különbségek eltűnésével. (10. táblázat)

10. táblázat Hoover-indexek értékei a lakosság száma és néhány más jellemző között különböző térfelosztások esetén, 1996, %

Dirichlet poligonok száma	Jellemző megnevezése		
	személygépkocsi	telefon	szja alap
2	13,5	16,3	13,1
3	14,3	18,5	14,1
4	14,4	17,7	16,6
5	15,1	19,4	18,2
10	17,4	24,1	20,4
15	18,7	27,0	21,1
20	19,1	28,7	22,0
25	19,8	30,6	23,2
30	20,8	31,7	24,3
50	22,1	34,1	26,1
60	22,2	35,0	26,6
80	22,7	35,6	27,0
100	23,3	36,2	27,9

Forrás: saját számítás a KSH és a PM adatai alapján

A területi adatok közötti korreláció azt mutatja meg, hogy két jellemző területi eloszlása mennyire hasonlít egymáshoz. A területi szintek növekedésével a korreláció szorosabbá válik, mivel a változók szórása csökken. (11-12. táblázat) A lehatárolás változtatásának hatására növekvő zónaszám mellett a korrelációs együttható csökkenése csak tendenciaszerű lesz, átmenetileg növekedhet is, hasonlóan viselkedik a maximális és minimális értékek hányadosához. Erre a 12. táblázat mindhárom adatsorozatban láthatunk példákat.

11. táblázat Az egy lakosra jutó személygépkocsi száma és néhány további jellemző közötti korreláció, 1996

Területi szint	Jellemző megnevezése		
	Telefonok száma	szja alap	Szociális segélyben részesítettek száma
Település	0,507	0,608	-0,176
Kistérség	0,687	0,646	-0,456
Megye	0,744	0,736	-0,566
Régió	0,894	0,907	-0,778

Forrás: saját számítás a PM adatai alapján

12. táblázat Egyes jellemzők súlyozott relatív szórása különböző térfelosztások esetén, %

Dirichlet poligonok száma	Jellemző megnevezése		
	telefonok száma	szja alap	Szociális segélyben részesítettek száma
3	0,829	0,973	-0,827
4	0,851	0,918	-0,810
5	0,849	0,912	-0,820
10	0,855	0,873	-0,760
15	0,750	0,749	-0,615
20	0,751	0,759	-0,630
25	0,790	0,782	-0,647
30	0,772	0,784	-0,603
50	0,767	0,744	-0,546
60	0,769	0,748	-0,558
80	0,746	0,719	-0,552
100	0,760	0,694	-0,580

Forrás: saját számítás a KSH és a PM adatai alapján

A részletesebb térfelosztás több információt szolgáltat a kevésbé részleteshez képest, mivel a területösszevonások az információk egy részének elvesztésével jár. Ez azonban nem jelenti azt, hogy az egyes mutatókat a legrészletesebb térfelosztás mellett kellene mindig kiszámolni. Lehetőség szerint olyan térfelosztás mellett kell dönteni, amelynek a vizsgálatát az adott probléma szempontjából legjobbnak tartunk. A területi szerkezet átfogóbb trendjeinek felderítéséhez például célszerűbb durvább területi felosztást használni, Magyarországon regionális vagy megyei szintű egységeket. Az egyes területi szintek eltérő viselkedése azzal a következménnyel is jár, hogy ha a területi különbségek időbeli alakulását kívánjuk nyomon követni, általánosságban nem jelenthetjük ki, hogy növekedtek, hasonlóak maradtak vagy csökkentek a különbségek, hanem csak adott területi szintre vonatkoztatva.

Végül fontos arra is utalni, hogy mindenféle módszer alkalmazása esetén óvatosan kell bánni a különböző terület egységekre vonatkozó mutatók összehasonlításával. Gyakran hasonlítják össze például az országok országon belüli fejlettségi vagy egyéb területi különbségeit, aminek jobb esetben csekély az információtartalma, rosszabb esetben pedig egyenesen megtévesztő is lehet. Ezek az összehasonlítások ugyanis mindkét fokozatú lehatárolási problémával terhelték: az egyes országok (vagy régiók) többnyire eltérő nagyságúak és a lehatárolási módjuk is nagyrészt különbözik, az aggregálás mértékében és módjában is.

Összegzés

A területi lehatárolások problémája a szűkebben vett területi kutatásokon túlmutató olyan alapvető statisztikai probléma, amit az elméletek alkotásánál, ellenőrzésénél és az adatok értékelésénél is figyelembe kell venni. Minden területileg aggregált adatot vagy területi átlagot használó elemzés során szembe kell nézni vele. Ennek tudatosodása azonban a kutatásokat végzők körében nem elégséges mértékű. A témával foglalkozó kutatók többsége a probléma viszonylagos mellőzöttségét általánosságban megemlíti, néhányuk viszont konkrét kutatásokat, cikkeket is megnevez, ahol a figyelmen kívül hagyása különösen indokolatlan. Így például Robinson 14 olyan, 1950-ben ismertebb tanulmányt sorol fel, ahol ökológiai korrelációt használtak egyéni viselkedés leírására, Fotheringham és Wong pedig három népszámlálási körzetek adatain nyugvó tanulmányt, ahol a módosítható egység problémájával egyáltalán nem számolnak. (Robinson, 1950; Fotheringham–Wong, 1991) Goodman három évvel Robinson cikkének megjelenése után írta, hogy 1950 óta számos ökológiai korreláción

alapuló következtetést felhasználó cikk jelent meg. (Goodman, 1953) Az azóta elmúlt fél évszázadban Robinson figyelmeztetése ellenére számuk nyilván tovább gyarapodott. A területi lehatárolás statisztikai következményeivel nem számoló kutatások egy része téves következtetéseket tartalmaz, más részének a szerzőik vélekedésénél korlátozottabb érvényűek a megállapításai, míg természetesen számos, bár területi adatokat felhasználó tanulmányban jellegükből adódóan nem játszik szerepet a kérdés.

A lehatárolási problémának nincs és nem is lehet matematikai, statisztikai megoldása. Jelentőségét általános érvennyel lehetetlen megállapítani, azonban néhány olyan körülményt fel lehet sorolni, ami általánosságban növeli a jelentőségét:

- a statisztikai módszer növekvő bonyolultságával együtt növekszik a jelentősége, minél bonyolultabb az alkalmazott statisztikai módszer, annál kiszámíthatatlanabb a hatása,
- a jellemzők gyakoriságának nagy változékonysága és a terület nagy heterogenitása esetén szintén növekszik a súlya,
- ha a kutatással konkrét döntéseket kell alátámasztani, növekszik a szerepe, még óvatosabban kell eljárni, ezen belül elsősorban a közösségi (állami, önkormányzati stb.) döntéseknél, és kevésbé a magánszféra döntéseinél,
- amennyiben több területegység szerkezetét, például egy jellemző területi egyenlőtlenségét kívánják összehasonlítani, a probléma nagysága hatványozódik.

A lehatárolási problémából az is következik, hogy a területi aggregátumokat használó modellek magyarázó ereje csak korlátozott lehet. A probléma egyetlen megnyugtató megoldásaként az elméletek és modellek megalkotásánál és az eredmények értékelésénél figyelemmel kell rá lenni.

Irodalomjegyzék

- Bach, L. (1981) The problem of aggregation and distance for analyses of accessibility and access opportunity in location-allocation models *Environment and Planning A* 13. 955-978. o.
- Duncan, O. D.–Davis, B. (1953) An Alternative to Ecological Correlation 665-666. o.
- Fotheringham, A. S.–Densham, P. J.–Curtis, A. (1995) The Zone Definition Problem in Location-Allocation Modeling *Geographical Analysis* 1. 60-77. o.
- Fotheringham, A. S.–Wong, D. W. S. (1991) The modifiable areal unit problem in multivariate statistical analysis *Environment and Planning A* 23. 1025-1044. o.
- Gehlke, C. E.–Biehl, K. (1934) Certain Effects of Grouping Upon the Size of the Correlation Coefficient in Census Tract Material *Proceedings of the American Statistical Journal New Series*, March 169-170. o.
- Goodman, L. A. (1953) Ecological Regression and Behavior of Individuals 663-664. o.
- Neprash, J. A. (1934) Some problems in the Correlation of Spatially Distributed Variables *Proceedings of the American Statistical Journal New Series*, March 167-168. o.
- Openshaw, S. (1977) A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling *Transactions of the Institute of British Geographers*, 2. 459-472. o.
- Openshaw, S. (1984) Ecological fallacies and the analysis of areal census data *Environment and Planning A* 16. 17-31. o.
- Openshaw, S.–Taylor, P. J. (1981) The modifiable areal unit problem in: *Quantitative geography: a British view* szerk: Wrigley, N.–Bennett, R. J. London, Boston, Henley Routledge and Kegan Paul 60-69. o.
- Robinson, W. S. (1950) Ecological Correlations and the Behavior of Individuals *American Sociological Review* 15. 351-356. o.

Sawicki, D. S. (1973) Studies of Aggregated Areal Data: Problems of Statistical Inference
Land Economics 49. 109-114. o.

Yule, G. U.–Kendall M. G. (1964) Bevezetés a statisztika elméletébe Budapest, Közgazdasági
és Jogi Könyvkiadó