2 ADATKEZELÉS, STATISZTIKAI ÉS SZÁMÍTÁSTECHNIKAI ALAPOK

2.1 Területi statisztika és térelemzés

A kutatási cél, a főbb vizsgálati témakörök (hipotézisek) meghatározása, a szükséges információk összegyűjtése után – gyakran már az információgyűjtéssel párhuzamosan – három fontos munkafázist hív elő a kutatási folyamat, még a tényleges elemzési szakasz előtt:

- az információk jellegének, méréselméleti tulajdonságainak s ebből következő *használhatóságának* ellenőrzése. Itt kerül először szembe a kutató a *statisztika* alapfogalmaival. Ezzel megelőzhetjük azt, hogy az elemzés közepén kerüljünk zsákutcába, s hogy netán adatainkat alkalmatlan módszerekkel dolgozzuk fel.
- az összegyűjtött információk *rendszerezése* (áttekinthető adattáblák előállítása) – ebben kiemelt szerepe van ma a *számítástechnika* eszközeinek és módszereinek.
- az összetettebb eljárások bevetése előtt fontos összefüggések tárhatók fel az alapinformációk (alapadatok) *egyszerű átalakításával is.* Ebben a fázisban az elemi *matematika* eljárásainak lehet szerepe.
- Számos területi vizsgálat induló lépése az információk *térképezése*, *grafikus ábrázolása.*

Ha van fejezete a kötetnek, amire mindenképp áll, hogy bármifajta teljesség igénye nélkül készült, akkor ez az, hisz akár a matematika, a statisztika vagy a számítástechnika alapfogalmainak áttekintése is önálló köteteket tölthetne ki. Ismételten hangsúlyozzuk, *ez a munka nem matematika-tankönyv, nem számítástechnikai sorvezető s nem statisztikai kézikönyv.* Ugyanakkor épp oktatási tapasztalataink alapján nem tűnt feleslegesnek a három tudományterület *néhány fontos fogalmának felidézése* itt is – mintegy emlékezetőként a legtöbbek által már (a középiskolában vagy a felsőfokú képzés speciális kurzusain) elsajátított ismeretekre. Itt ismertetjük a területi kutatások kvantitatív módszerei kapcsán leggyakrabban használt s a kötetben is alkalmazott *jelöléseket, formulákat is. E fejezet tartalmazza az ELTE geográfusképzés alapozó matematikai, statisztikai és számítástechnikai kurzusain sorra vett legfontosabb kérdésköröket.* Ezek oktatási anyagokba, jegyzetekbe foglalásában a hazai műhelyek közül kétségkívül a szegedi egyetem oktatói állnak az élen, akik tollából vagy szerkesztésében a korábbi munkák után a kilencvenes években is több kötet látott napvilágot (*Koppány Gy. et al 1995, Herendi I. et al. 1996, Abonyiné Palotás J. 1999, Katona T. - Lengyel I. szerk. 1999*). A bevezető egyetemi kurzusokon jól használható összefoglalója a társadalomföldrajzban használható elemi matematikai eljárásoknak Vuics Tibor feladatgyűjteménye (*Vuics T. szerk. é.n.*) A matematikai-statisztika kézikönyvei közül *Hunyadi L. – Vita L. 2002* munkája a legújabb, a sokáig használt klasszikusok közül kiemelhető *Köves P. – Párniczky G. 1981*. Mindezek haszonnal forgathatók. Mivel a matematikai és a statisztikai ismeretek nem avulnak el, bármely régebbi átfogó könyv használható ma is. Mindez kevésbé mondható el a számítástechnikára, ott különösen érdemes figyelni a legújabb kiadványokra.

A kvantitatív elemzéseket a területi kutatásokban a vizsgálat térbelisége és az elemzett egységek, objektumok jellege szerint két nagy csoportba oszthatjuk:

- A *területegységek* (települések, régiók, országok vagy általában bármilyen konkrét térfelosztás eredményeként kapott területegységek rendszereinek) vizsgálata a klasszikus területi adatmátrixból (→ **2.2.2**) indul, s elemzési eszköztára az *általános és területi statisztika* (*spatial statistics*) módszereire támaszkodik[1]. A térbeli, területi azonosítás ezekben az esetekben a megfigyelési egységek nevével, térképi ábrázolásával történik. E vizsgálatok számítástechnikai bázisát a hagyományos táblázatkezelő és statisztikai, valamint térképrajzoló programok jelentik.

---

[1] Ezt az ismeretkört fogja át a hazai regionális elemző szakirodalom immár klasszikusnak számító munkája, a *Sikos T. Tamás* által szerkesztett kötet *(1984),* s hasonló szemléletű tanulmányokat tartalmaz *Kulcsár V. szerk. 1976,* amely az első „modern” magyar nyelvű kvantitatív elemző tanulmánykötetnek tekinthető.

Ebben a szemléletben *a* regionális és települési elemzések beilleszkednek *a társadalomelemzés* más dimenziói sorába (*Bukodi E. 2001*). Ilyen jellegzetes közelítést képez *a demográfiai* (nem, életkori csoportok, családtípusok szerinti) vizsgálat; *a társadalmiosztály-dimenzió* (foglalkozási csoportok, munkaerő-piaci szektorok, ágazatok szerinti megoszlás); különböző *társadalmilag „veszélyeztetett" csoportok* (képzetlenek, tartós munkanélküliek, fiatal munkanélküliek, fogyatékosok stb.) vizsgálata.

- A térben *helyzetparaméterekkel (helykoordinátákkal) azonosított* adatrendszerek vizsgálatakor a térbeliség, a szomszédsági relációk, a távolságeloszlások, a konfiguráció vizsgálata a *térelemzés* (*spatial data analysis, exploratory spatial data analysis*) eszközeit igényli A témakörben számos átfogó módszertani munka látott már napvilágot (az újabbak közül: *Fotheringham, A. S. et al. 2000*). A számítógépi munkában itt már speciális programokra van szükség. Ez az elemzési szemlélet általánosabb kérdéskörökhöz nyit utat, mint a sajátos területi statisztikai közelítés, s így a regionális elemzéstől tartalmilag távoli, generális problémakörökben is felbukkan (ilyen kérdéskör például a *képfelismerés*) – analógiákat kínálva a társadalmi folyamatok térbeliségének elemzéséhez. (A legújabb hazai szakirodalomban ilyen tematikát is tartalmaz *Dusek T.* már többször idézett értekezése, valamint *Varga A. 2002* térökonometriai áttekintése.)

A térelemzés (ezen belül is a térökonometria) egyik vezető nemzetközi szaktekintélye, *Luc Anselin* a döntően GIS alapú társadalmi térelemzés alábbi kulcsterületeit és hatásait jelöli meg (*Anselin, L. 1999):*

- *adatintegráció* (különböző tartalmú adatbázisok együttes elemzése a térbeli lokalizáció segítségével, egyazon területre vonatkozóan),
- a térbeli eloszlások, konfigurációk feltárását segítő *vizualizáció,*
- empirikus *térstatisztikai és tér-ökonometriai* elemzések,
- *a térbeli (földrajzi) gondolkodás erősödése* a társadalomkutatásban (lásd például Krugman „új gazdasági földrajzát) és a döntéshozatalban,
- a *társadalmi tér* szemléletének alakítása különböző tudományterületeken,
- a *tér-idő relációk* kutatásának előtérbe kerülése.

Mindezek fontos feltételeként hangsúlyozza a térelemzési ismeretek integrálását az oktatásba, képzésbe. (Hasonló elemzési momentumokat emel ki *Bivand, R. 1998*.)

Érdemes még megemlíteni, hogy a módszertani irodalomban találkozhatunk a *térstatisztika* kifejezéssel is, ez azonban nem a helyzetparaméterekhez kötődő, geokódolt adatok elemzését jelenti, hanem a sokváltozós, n-dimenziós adatrendszerek („terek") vizsgálati módszereire utal (*Füstös – Meszéna – Simonné 1997*). Ezek közül kötetünkben részletesebben a faktoranalízissel → **5.5** foglakozunk[2].

A kétfajta közelítés rokonságát és különbségét jól érzékeltethetjük, ha néhány alapvető fogalom, illetve kvantitatív eljárás szempontjából szembesítjük őket (*2.1. táblázat*).

[2] Rokon-módszer a regionális vizsgálatokban ritkábban használt többdimenziós (vagy sokdimenziós) skálázás (*Lengyel I. 1999*) vagy a diszkriminancia-analízis (*Obádovics Cs. 2004*). E két eljárás inkább az *ordinális változókkal, hasonlósági mértékekkel* operáló véleménykutatási, szociológiai vizsgálatokban gyakori. A szűkebb értelemben vett regionális elemzések ellenben általában magasabb mérési szintű változóra épülnek, amelyeket ritkán indokolt méréselméleti szempontból „lebutítani".

Fogalomkör	Területi statisztika	Térelemzés
Azonosítók	Nevek	Helykoordináták, vektorok
Középértékek	Átlag	Térbeli közép (súlypont)
Differenciáltság	Szórás	Standard távolság
Megoszlások	A jellemzők eloszlásfüggvényei	Térbeli alakzat, konfiguráció
Kapcsolatok	Korreláció, regresszió	Területi autokorreláció és –regresszió

2.1. táblázat *A területi statisztika és a térelemzés jellegadó fogalompárjai*

A területi statisztika és a térelemzés természetesen össze is kapcsolódhat. A területi adatok összetettebb matematikai-statisztikai elemzésének maga az egységek térbelisége szab korlátokat, hisz ezek a megfigyelési egységek nem tekinthetők egy véletlen minta elemeinek, mert épp a térbeli közelség, szomszédság következtében egymásra is hatnak, s így jellemzőik között kapcsolat, *területi autokorreláció* (→ **4.3)** léphet fel.

Más kérdés azonban az, hogy egy konkrét, empirikus adatsor – mondjuk a GDP-nek a világ országaira vonatkozó keresztmetszeti adatsora – egyáltalában valamely matematikailag jól definiált elméleti eloszlás „mintájaként” vizsgálandó-e vagy sem. Ez a matematikai-statisztikai eszközök empirikus használatának egyik máig nyitott vitakérdése.

A területi elemző legtermészetesebb szempontja (mondhatnánk axiómája) az, hogy térbeli, területegységekhez rendelt (településekre, körzetekre, országokra vonatkozó) információkkal dolgozik. Ez annyira magától értetődő, hogy sem maguk az információk, sem a velük kapcsolatos elemzési eljárások kapcsán ezt a jellemzőt külön általában nem is taglalják. Márpedig a lokalizáltság, a helyhez kötés, a térbeli kiterjedés és érvényesség kapcsán nem árt néhány feltételt megvizsgálni, mielőtt bonyolult elemzésekbe kezdenénk. Vegyünk néhány ilyen szempontot:

- Az adott jelenség *valóságos szerveződési egységeiben és szintjein* vizsgálódunk-e, vagy épp ott, ahol, amire az információk rendelkezésre állnak?
- Miként módosítja a vizsgált összefüggéseket, ha *megváltoztatjuk a térségi kereteket és szinteket* (aggregáljuk vagy dezaggregáljuk az adatokat)?
- Mit kezdjünk a térben megosztott, *egyetlen helyhez vagy téregységhez nem vagy csak részben köthető* jelenségekkel, milyen térbeli keretekben elemezzük őket?
- Milyen *elemi* (tovább nem bontható illetve tovább nem bontott) egységei vannak a területi vizsgálatoknak?
- Elég *egyetlen térségi szinten vizsgálódni*, netán követelmény minden jelenség esetében az, hogy különböző aggregációkban is elvégezzük az elemzést?
- Milyen feltételekkel használhatók a területi elemzésekben *a geometria jól ismert térelemei* (pontok, vonalak, térrészek), mint sajátos lokalizációs modellek?
- Milyen *gyakorlati* gondokat okoznak mindezen kérdések a (területi) *társadalomirányításban*?

2.2 Műveletek területi adattáblákban és adattáblákkal

2.2.1 Szám, vektor, mátrix

A területi elemzések információbázisai, adattáblái leggyakrabban számokból (skalárokból), s az ezek sajátos elrendezésével létrejövő vektorokból és mátrixokból állnak. Először néhány ezekkel kapcsolatos matematikai összefüggést idézünk fel.

Szám

A szám a matematika legalapvetőbb, önállóan nem definiált fogalma. Megkülönböztethető dolgok *megszámlálásával* jutunk *a természetes számokhoz* (1, 2, 3…). Ezekből minden más szám – különböző matematikai műveletek felhasználásával – levezethető.

Ilyen levezetett szám a 0 (nulla) is, amely tehát nem tartozik a természetes számok közé, hanem már művelet eredménye, két azonos szám különbségéből adódik (a-a = 0). A nullánál kisebb számokat negatív, a nagyobbakat pozitív számoknak nevezzük. Két különböző előjelű szám is lehet azonos *abszolút értékű,* ha hányadosuk -1 (*a/b*=-1, például 5 és -5). A számok abszolút értékének önálló jelölése van: *a* abszolút értékét $|a|$-val (függőleges vonalak, nem szokásos zárójelek!) jelöljük, ennek megfelelően: $|5|=|-5|=5$.

Kitüntetett tulajdonságú szám az 1. Bármely számot 1-gyel szorozva önmagát kapjuk eredményül. A számok tetszőlegesen nagyok illetve kicsik lehetnek, mínusz és plusz végtelen között oszlanak el (jelölés: $-\infty$ és $+\infty$).

Kerekítés, normál alak

A valós számokat általában végtelen sok számjeggyel adhatjuk meg. Ha a szám ilyen előállításában valamely adott értéknél kisebb helyiértékű számjegyeket olyan módon hagyjuk el, hogy a megmaradó számjegyekkel leírt számot az elhagyott számjegyek nagyságától függően, egy adott előírás szerint módosítjuk, *kerekítésről* beszélünk. Tízes számrendszerben úgy kerekítünk, hogy ha a legnagyobb helyiértékű elhagyott számjegy

5-nél kisebb, akkor a megmaradó számokat változatlanul hagyjuk
5-nél nagyobb, akkor a legkisebb helyi értékű el nem hagyott számjegyet 1-gyel növeljük
épp 5, akkor csak abban az esetben *nem* növeljük a legkisebb helyi értékű el nem hagyott számjegyet, ha az páros és az elhagyott szám után csupa nulla áll (az alábbi számpélda második esete ez, mivel viszonylag ritkán fordul elő, „durva" szabályként azt mondhatjuk, hogy az *5-öt is felfelé kerekítjük*.)

Szám:	8,72**5**365	Kerekítve:	8,7**3**
	8,7**25000**		8,7**2**
	8,7**15000**		8,7**2**

Az elemzések során, tartalmi szempontból akkor indokolt kerekíteni, ha az adott jelzőszám esetében az elhagyott értékeknek nincs érdemi jelentőségük, a kerekítési hiba hatása ezért elhanyagolható. Kis elemszámú mintákból számított korrelációnál (**→4.2**) sem indokolt a „tudományosnak" látszó sokértékű együttható közlése és részletező interpretációja: megyei szintű rangkorreláció 0,3915 és 0,3500 értéke egyaránt 0,4-es korrelációt jelez. (A témakör további matematikai finomságairól lásd *a kerekítés-elmélet* szakirodalmát.)

A számok alakjával összefüggő átalakítás a nagy számok ún. *normálalakban* történő (helytakarékos) megadása. Egy szám normál alakja az egy egész helyiértékű alak és 10 megfelelő hatványának szorzata. Például: 2002. január 1-én az ország népessége 10 174 853 fő volt, ennek normálalakban megfelel az $1{,}0174853*10^7$ kifejezés, aminek szokványos (például a legtöbb táblázatkezelő programban használt) jelölése: 1,0174853E + 07. (Az más kérdés, hogy a számbavételi nehézségek miatt legfeljebb ezres nagyságrendben megbízható a lélekszám.)

A változók jelölése

Matematikai összefüggésekben, képletekben, egyenletekben jellemzően nem egyes, konkrét számok, hanem általában – az adott összefüggésben megengedett, értelmezhető – számok halmazai, csoportjai, a *változók* szerepelnek (a statisztikában ez a különbség jelenik meg az *ismérv* és az *ismérvérték* fogalompárban). Ezek jelölésére különböző (latin, görög, héber) ábécékből vett betűket használnak. Bár bizonyos matematikai területeken e betűjelölések jellegzetesek – a geometriában gyakoriak a görög betűk (α,β), az algebra a latin betűsor elejéről válogat (a,b,c), míg a statisztika ugyanezen betűsor végéről (x, y, z) – hangsúlyozni kell, hogy ezek nem kötelezően használandók, csak kényelmes *konvenciók. Egy új, ismeretlen képlettel találkozva nem a betűkre, hanem az összefüggés tartalmára, jelentésére kell figyelni!*

Vektor

A vektor a számfogalom kiterjesztése. *A rendezett szám-n-eseket vektornak nevezzük.*

$$\underline{V} = (a_1, a_2, ...a_n)$$

A rendezettség a vektorban szereplő számok rögzített sorrendjét jelenti, azonos számjegyeket, de különböző sorrendben tartalmazó vektorok különbözőek, például pl. A (4,8,0,6,2,2,) és B (2,2,6,0,8,4) két különböző vektor. Két vektor akkor egyenlő egymással, ha ugyannyi számból áll (másként: a *dimenziójuk* azonos) s ugyanazon (az első, második, n-edik) helyen ugyanazon érték áll mindkét vektorban. A vektort mind sorokban (sorvektor), mind oszlopokban (oszlopvektor) elrendezett számok alkothatják. A vektor a geometriában irányított szakaszokkal reprezentálható.

Azonos dimenziószámú vektorokkal műveletek végezhetők. 0-vektornak (nullvektornak) nevezzük azokat a vektorokat, amelyek minden eleme 0. Ezt két azonos vektor különbségeként kaphatjuk, úgy, hogy a megfelelő helyen álló számokat kivonjuk egymásból. (Végtelen számú 0 vektor lehetséges, hisz minden dimenziószámhoz különböző 0-vektor rendelhető).

Vektorok esetében kétféle szorzási művelet is értelmezhető, ezek közül a skalárszorzat kerül gyakrabban elő a területi elemzésekben. Ez $V(v_1, v_2 \ldots v_n)$ és $W(w_1, w_2 \ldots w_n)$ azonos dimenziószámú vektorok esetén értelmezhető, s a két vektor azonos helyen álló elemeinek szorzatösszegeként egy *számot* (skalárt) ad eredményül:

$$b = \sum_{1=1}^{n} v_i w_i$$

Ugyancsak azonos dimenziószámú vektorokra értelmezett az ún. *vektoriális szorzat*, amely egy *vektort* ad eredményül, amelynek első eleme a két vektor első helyen álló elemének szorzata, a második a második helyen álló elemeké és így tovább:

$$V*W = T\,(v_1w_1, v_2w_2 \ldots v_nw_n)$$

Koordinátarendszerek

A koordináta (vagy vektor-) geometria a matematika többi nagy területét (algebra, analízis) kapcsolja össze a geometriával. A koordinátarendszerek rendezett szám n-esek (vektorok) és az euklideszi tér pontjai között teremtenek megfeleltetést.

A területi elemzésben is leggyakrabban a *Descartes-féle derékszögű koordinátarendszert* használják rendezett számpárok és a sík pontjainak megfeleltetésére (súlypont → **2.6**, regresszió-számítás → **4.4**)

A koordinátarendszer kezdőpontja, origója tetszőlegesen választható meg. A tengelyek eltolása (s a ritkábban előforduló elforgatása) – ez az ún. tengely-transzformáció, amelynek során az új tengelyek metszőpontját tekintjük origónak – *az eredeti pontok koordinátái megváltoznak, de ez nem változtatja meg a pontok egymáshoz viszonyított helyzetét és távolságát.* Ha a koordinátarendszer origóját a C (a;b) pontba toljuk el, akkor az új rendszerben az A (x; y) pont új koordinátái x' = x-a illetve y' =y-b lesznek.

Magyarországi vizsgálatok esetében, amikor az elemzés alappontjai az ország települései (városai), előnyös a koordinátarendszert úgy kijelölni, hogy annak origója a főváros középpontjában legyen, ekkor ugyanis a pontok (települések) helykoordinátáinak előjeleiből azonnal következtetni lehet azok földrajzi pozíciójára is. A számítási eredmények (pl. súlypontok) azonban akkor sem mutathatnak más térbeli konfigurációt, ha a koordinátarendszer középpontja a térkép bal alsó sarkában van (ilyenkor minden pont mindkét koordinátája pozitív).

A koordinátarendszerben elhelyezett pontok vizuális értelmezésével nagyon óvatosan kell bánni, hisz azt a koordinátarendszer skálája (egységei) jelentősen befolyásolhatják → **2.9.**. Függvények, statisztikai eloszlások egyszerű grafikus ábrázolásakor a két tengelyen ugyanis nem feltétlenül azonosak az egységek (ez gyakran célszerű és megengedhető). A pontok távolságát nem vonalzóval kell mérni az ábrán, hanem a megfelelő távolságfüggvény segítségével.

A földrajz legismertebb koordináta-rendszere a földi szélességi és hosszúsági hálózat, ami egy gömbi polár-koordinátarendszer, hisz ott minden ponthoz egy *távolságérték* (a Föld elméleti középpontjától mérve) és a két *irányszög* (a greenwichi délkörhöz illetve az Egyenlítőhöz viszonyítva) határozza meg a pontok helyzetét. A földi gömbi koordináták használata a társadalmi jelenségek vizsgálatakor viszonylag ritka, bár nagyobb térségek elemzésekor előfordul. Kisebb terekben végzett elemzések esetében sík területet feltételezünk a földi térben lokalizált elemek helyzetének meghatározásakor. Speciális koordinátarendszer az ún. *háromszögdiagram* → **2.9.7**.

Mátrix

A mátrix ugyancsak a számfogalom kiterjesztése. Ha számokat n sorból és m oszlopból álló *táblázatokba* rendezünk, kapjuk a mátrixokat. (Általánosságban a matematika nem csak számok, hanem bármilyen matematikai kifejezés ilyen táblázatos elrendezését mátrixnak nevezi.) Az egy sorból illetve egy oszlopból álló mátrix vektorként is értelmezhető.

A matematika mátrix fogalma – egyebek mellett – abban különbözik a táblázat fogalmától, hogy abban a sorok és az oszlopok *tartalmának* nincs jelentősége – *a mátrix oldal- és fejléc nélküli táblázat. Ebben az értelemben a területi kutatás jellemzően csak táblázatokkal dolgozik, még ha mátrixnak hívja is őket*.

A mátrixokat leggyakrabban nagy latin betűkkel (**A, B, C**) vagy ún. általános (az i.-ik sorban és a j.-ik oszlopban található – elemükkel jelölik: $\{\mathbf{x_{ij}}\}$. Ha a mátrix méretét – sorainak és oszlopainak számát – is fontos megadni, akkor a fenti jelölések egy n sorból és m oszlopból álló mátrix esetében így módosulnak: ${}_{\mathbf{n}}\mathbf{A}_{\mathbf{m}}$ illetve ${}_{\mathbf{n}}\{\mathbf{x_{ij}}\}_{\mathbf{m}}$.

Mind a vektorokkkal, mind a mátrixokkal a számokhoz hasonló műveletek (összeadás, szorzás) végezhetők, speciális szabályok szerint. Gyakran előforduló művelet *mátrix és vektor szorzata*, amivel gyakran találkozhat a területi elemző is, összetettebb statisztikai módszereket használó publikációk összevont módszerleírásaiban. Ez a művelet az ún. *sor-oszlop* skalárszorzás. *Az **y** vektor elemeit az **A** mátrix soraiban található elemekkel rendre összeszorozzuk, majd ezeket a szorzatokat összeadva egy új vektort kapunk eredményül*. A szorzás feltétele, hogy a mátrixnak annyi sorból kell állnia, ahány elemű a műveletben szereplő oszlopvektor (ezt érzékelteti az alábbi mátrix-egyenlet).

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0{,}5 & 0{,}5 \\ 0 & 0 & 0{,}5 & 0 & 0{,}5 \\ 0 & 0 & 0{,}5 & 0{,}5 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix} = \begin{pmatrix} y_2 \\ y_1 \\ 0{,}5y_4 + 0{,}5y_5 \\ 0{,}5y_3 + 0{,}5y_5 \\ 0{,}5y_3 + 0{,}5y_4 \end{pmatrix}$$

Jellegzetes jelölése e műveleteknek: **A** $\underline{\mathbf{y}}$ (**A** mátrixot szorozzuk $\underline{\mathbf{y}}$ vektorral). Nem megrémülni, megérteni (rövid matematikai utánképzést igényel csak)! Legtöbbször nagyon egyszerű műveletekről van szó, összevont jelöléssel, az alapadatok ügyes, mátrix és vektor formájú elrendezését felhasználva.

2.2.2 A területi adatmátrix

Amikor rendelkezésre állnak használható számszerű információk, akkor a *területi* elemzésekben a tér leggyakrabban *a megfigyelési egységekben* van jelen. Országokra, régiókra, településekre vonatkozó jellemzők vizsgálata folyik. Ezeket – a területi statisztikai kiadványokat megtöltő – adattáblákat nevezzük *területi (földrajzi) adatmátrixoknak (2.1. ábra)*. A mátrix sorai egy-egy területegység (területi, települési megfigyelési egység) különböző jellemzőit tartalmazzák, oszlopai pedig egy-egy jellemző értékeit a területegységekben.

$$\begin{array}{cc} & \text{Jellemzők} \\ \text{Területegységek} & \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{im} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nj} & \dots & a_{nm} \end{bmatrix} \end{array}$$

2.1. ábra *Területi adatmátrix*

Míg a terület, mint megfigyelési egység tehát közös jegy, a jellemzők (gazdaságföldrajzi, regionális gazdaságtani, településszociológiai, demográfiai, politikai vagy más indikátorok) tartalma az, ami az egyes területi tudományokat elválasztja.

Az oszlopok és a sorok szerepe felcserélhető. Az ilyen elrendezést azonban ritkábban használjuk (leginkább akkor, ha néhány területegység, ország közvetlen összehasonlítása a cél). A legalapvetőbb statisztikai mutatók számításakor is kedvezőbb, ha a változók szerepelnek az oszlopokban, s több összetettebb módszer esetében (faktoranalízis **→5.5)** követelmény az is, hogy a megfigyelési egységek (területegységek) száma jelentősen haladja meg a jellemzők, változók számát, így az ajánlott sor-oszlop elrendezés kényelmesebb, áttekinthetőbb.

A területi adatmátrix egyes elemei (a számítástechnikában: *cellák*), sorai, oszlopai, résztömbjei szerint vehető elemzésbe, egyszersmind sajátos vizsgálati utakat teremtve:

- a *sorok* (a számítástechnikában: *rekordok*) szerinti elemzés az egyes területegységek jellemzésére, a sorok *összehasonlítása* egymással a megfigyelési egységek belső térszerkezeti összetételének, struktúrájának összevetésére ad lehetőséget. Ha mátrixunk csak egyetlen sorból áll, akkor *sor-vektornak* is nevezhetjük;
- az *oszlopok* (a számítástechnikában: *mezők*) vizsgálata az adott jelenség térbeli eloszlásának (térképezésének), a területi differenciáltság mérésének kiindulópontja, az oszlopok *összehasonlítása* a különböző társadalmi szférák vagy jelzőszámok közötti területi kapcsolatok (asszociáció, korreláció), összefüggés, együttmozgás vagy elkülönülés vizsgálati útja. Ha mátrixunk csak egyetlen oszlopból áll, akkor *oszlop-vektornak* is nevezhetjük.

A nagy adatmátrixokat, vagy több, egymással összekapcsolt adatmátrix együttesét *adatbázisnak* nevezzük. Mindenkor érdemes arra is vigyázni, hogy a különböző jelzőszámokat azon a térségi szinten, területi keretben használjuk, ahol tartalmuk valóságosan megfogható, releváns. Nincs értelme a fajlagos acélfelhasználást a fővárosi kerületek szintjén vizsgálni, még ha – nagy nehézség árán – talán meg lehetne valamilyen számértéket erre is becsülni, s ugyanígy eléggé formális a szociális segélyek nagytérségi (regionális) elemzése, hisz ez az egyének, családok, kisebb társadalmi aggregátumok szintjéhez kötött. Mindez kissé általánosítva azt jelenti, hogy a területi adatmátrix sorai és oszlopai (mivel azok tartalma lényeges) nem kombinálhatók véletlenszerűen.

Az ismérvek rövid elnevezése

Egy rendkívül gyakorlatias, de mindenkor felbukkanó feladat, hogy az elemzésbe vett, bonyolult módszerekkel feldolgozásra váró adatsorokat (a területi adatmátrixok oszlopfejeiben, fejlécében, vagy a köztük lévő kapcsolatot leíró egyenletben) rövid jelöléssel is azonosítani kell. (Egy többváltozós regressziós analízis leírásába és az eredmények közlésébe aligha illeszthetők be teljes elnevezésükkel például olyan változók, mint az *egy főre jutó GDP 1999. évi* adata vagy *a városi népesség aránya 2000.)* Arra, hogy miként rövidítsük, jelöljük az adatot, nincs általános előírás. A magam „módszerében” az alábbi szempontokat szoktam érvényesíteni:

- A rövidítés valóban legyen rövid (8, legfeljebb 10 karakter)
- A karakterek között ne legyenek hosszú ékezetes magyar betűk (ezt sok számítógépi program nem szereti)
- A rövidítés utaljon a változó tartalmára
- Szerepeljen benne az időpont is (ez főként akkor indokolt, ha az adattáblában különböző időpontra vonatkozó adatok szerepelnek).

Ennek megfelelően a fenti változókat így jelölném: *GDPPOP99* illetve *VARPOP00*. Aki érti (például azt, hogy a POP betűhármas miért szerepel mindkét mutatóban, és mire utal) és akinek tetszik a megoldás, válassza ezt, akinek jobb, szebb megoldása van, tegyen úgy, az ajánlott szempontokat nem teljesen félretéve!

2.2.3 Műveletek adatsorokkal

Új adatok előállítása több adatsorból

A területi elemzések során nagyon gyakori, hogy több adatsorunk is van, s nem pusztán egyenként vizsgáljuk őket, hanem (ha lehetséges) egyszerű számtani műveleteket is elvégzünk velük. Itt a területi adatmátrix oszlopaival (jellemzők, adatok) végzett műveletekről van szó, a sorok (a területi megfigyelési egységek) változatlanul maradnak.

Összeadással kapott új adatsorok

Az adatsorok egyszerű összeadása tartalmilag homogén adatok esetében kerül szóba. Ezen az úton a *több részelemből álló aggregátumok* egészére kaphatunk adatokat (az egyes ágazatokban foglalkoztatottak területi adatainak összeadása például a teljes foglalkoztatotti kört, az egyes korcsoportoké pedig az össznépességet adja ki.)

Azt, hogy milyen adatokat tekintünk *tartalmilag homogénnek* nem csak maga a dolog vagy jelenség, hanem az értékelési szempont is befolyásolja. A búza, a napraforgó és a burgonya termésmennyiségét a mezőgazdaságot elemezve nem szokás közvetlenül összeadni, de a vasúti áruszállítás szempontjából ezek tömege a mérvadó, s ekkor már összeadhatók. (A naturálisan különböző dolgok „közös nevezőre” hozását, amikor tehát valami módon mégis „összeadjuk” őket, külön tárgyaljuk → **5.4.1**).

Jellemző művelet az egy-egy hosszabb időszak területi folyamatainak elemzésekor *a természetes egységekben* (naturáliákban) mért *területi idősorok* egyedi időpontokra (általában egyes évekre) adott értékeinek összeadása is az időszakra egészére. Gyakori példája ennek a demográfiai mozgalom éves adatainak, vagy a természetes egységekben kifejezett fejlesztések – pl. az éves lakásépítés, az épített autópályák hossza – volumenének összegzése, kumulálása. A növénytermelési adatok esetében az időjárás változásával összefüggő, ismétlődő, de időben véletlenszerű termésingadozások hatásának kiszűrésére általában több (3-4) év adatának ún. mozgó átlagával számol az összehasonlító agrárstatisztika.

Kivonással kapott új adatsorok

Ha túllépünk azon a logikailag és számtanilag természetes összefüggésen, hogy, ami összeadható, az ki is vonható (például a területi idősorok egyes adatoszlopait kivonva egymásból az éves relatív többletek és hiányok tárulnak fel), e példa mellé újabbak is felsorakoztathatók.

Szemben az összeadással, ami kommutatív (felcserélhető), azaz $A+B = B+A$, a kivonás nem, azaz $A - B \neq B - A$. S nem áll a fenti kijelentés ellentettje sem, miszerint, ha két adat egymásból való kivonásának van értelme, akkor az összeadásának is van: ha az ország teljes népességéből kivonjuk a városlakók számát, megkapjuk a községekben lakók lélekszámát, ha azonban összeadjuk a teljes népességet és a városlakókat, akkor (millióban számolva Magyarországon) egy megjátszható lottószám az eredmény, semmi más.

Vegyünk azonban még érdekesebb példát a kivonással kapható új adatokra. Erre jó terep a hazai területi kutatások egyik leggyakrabban használt területi kincsestára a személyijövedelem-adó adatbázisa. Ebben az APEH 1988 óta lényegében minden évre hozzáférhetővé tesz az ország minden településére négy fontos adatot:

az adóköteles jövedelem (*J*, forintban)
a személyi jövedelemadót fizetők száma (*A*, főben)
a befizetett személyijövedelem-adó (*S*, forintban)
az állandó népesség (*N*, főben)

Ezt az adatbázist használva, teljesen új minőségű adatot kapunk például akkor, ha az adóköteles jövedelemből (J) kivonjuk a befizetett személyi jövedelemadót (S), ekkor áll ugyanis elő a $J–S= N$, nettó jövedelem.

Szorzással kapott új adatsorok

A szorzás kevésbé gyakori művelet új adatok előállítására a területi elemzésben, bár itt is előfordul, leggyakrabban az árakkal, költségekkel kapcsolatos elemzésekben. Az utazási vagy szállítási költség például a megtett út és a tarifák szorzata, de hasonló mutató az üzemanyag fogyasztás is.

A szorzás, mint alapvető matematikai művelet természetesen nagyon gyakori minden más összetettebb matematikai formula, egyenlet esetében.

Osztással kapott új adatsorok

A műveletek közül talán az osztás a legnagyobb új adattermelő. Osztással kapjuk ugyanis az alapadatokból a *relatív vagy fajlagos* adatokat. A fenti adatbázisból véve a példát: *J/N* adja az egy főre (lakosra) jutó adóköteles jövedelmet, *A/N* pedig az adózók hányadát az összes népességből (ha ezt a hányadost 100-zal szorozzuk, akkor az arányt százalékban kapjuk).

A területi vizsgálatokban leggyakrabban a *népesség egészére* (illetve speciális jellemzők esetében az *érintett népességi* körre: a fenti adatok közül ilyen az adófizetők száma), a *területnagyságra* vetített mutatószámokat használnak. Előbbiek a *fejlettség, ellátottság, hatékonyság* jellegzetes mutatóit adják, utóbbiak pedig a különböző területi *sűrűségi, koncentrációs* mutatókat (a legismertebb a népsűrűség, amely a lélekszám és a terület hányadosa).

A gazdasági fejlettségi és sűrűségi mutatók között egyszerű összefüggés van, a gazdasági sűrűséget a gazdasági fejlettség és a népsűrűség szorzatára bonthatjuk.

GDP/terület = (GDP/népesség)*(népesség/terület)

azaz:
gazdasági sűrűség = (gazdasági fejlettség)*(népsűrűség)

Új fajlagosok konstruálásával érdemes óvatosnak lenni. Formálisan számítható ugyan, de a vasútvonalak egy óvónőre vetített hosszának kistérségi különbségeit vélhetően csak nagyon cizellált kutatási problémakörben érdemes önálló indikátorként elemezni. Mindez arra int: törekedjünk érdekes, új mutatók kiszámítására, de értően, a tartalomra figyelemmel!

Gyakori az, hogy bizonyos fajlagos adatok és azok *reciproka* egyaránt értelemmel bír. A fenti adatokból számítható *A/N* (adózóarány) reciproka, az *N/A* mutató azt fejezi ki, hogy egy adózónak hány embert kell eltartania. Mindez bizonyos esetekben jól kihasználható. Lehet olyan vizsgálat, amikor a várossűrűség (városok/terület) máskor ellenben ennek épp a reciproka (terület/városok) a jó mutató. Előbbi inkább a területi fejlettség, utóbbi az intézményi ellátottsági szükségletek tervezésének témakörében kerül szóba.

Fejlettségi vizsgálatok során gyakran használják az elemzők a munkanélküliségi rátát (munkanélküliek és az aktív keresők aránya, százalékban kifejezve). Ennek azonban az a sajátossága, hogy „kilóg” a hagyományos fejlettségi, jóléti mutatók köréből, hiszen minél nagyobb az értéke, annál rosszabb a helyzet az adott térségben. Itt is szóba jöhetne a reciprok-mutató (aktívak/munkanélküliek), ennek azonban elég nehezen tudunk értelmet adni. Ebben az esetben a munkával rendelkezők aránya (100-munkanélküliségi ráta) lehet az a jelzőszám, aminek nagy értékei a kedvező, alacsony értékei a kritikus térségi állapotokra utalnak.

A fajlagosok mértékegységei

Az adatok értelmezéséhez fontos azok mértékegységének, dimenziójának ismerete. Osztással kapott, fajlagos adatok esetében az új mértékegység a számlálóban illetve a nevezőben szereplő adat mértékegységének hányadosa: népsűrűség = népesség/terület, ezért mértékegysége fő/km^2. A gazdasági fejlettség mérésére használt jövedelemfajlagosok mértékegysége pénznem/fő. De ha konkrét mutatót emlegetünk (térképezünk), például a gyakorta használt egy főre jutó GDP-t, elégséges a pénznem megnevezése is. Számos olyan jelzőszám van, aminek a számlálója és a nevezője azonos mértékegységű, így az osztás után ezek mértékegység-nélkülivé, dimenziótlanná válnak. Ezt oldja fel az a megoldás, hogy a számlálót a nevező százalékában adjuk meg (ilyen a városi népesség aránya %-ban). Vannak nem szokványos mértékegységű mutatók is, így például a vasúthálózat sűrűségének dimenziója 1/km (a hossz/terület hányados következményeként). Összetettebb modellek formuláiban ugyancsak sajátos mértékegységek szerepelnek (pl. a potenciálmodellben → **6.7.2** dollár/km lehet a dimenzió).

Részben a mértékegységgel függ össze az, hogy bizonyos fajlagosok esetében a számláló illetve a nevező maga milyen mértékegységű, nagyságrendű legyen. Vegyünk néhány példát. Bár a *népsűrűség* számításánál egyaránt értelmes, ha a népességet egy négyzetkilométerre vagy egy hektárra vetítjük, a szokásos eset az egy négyzetkilométerre vetítés (ennek ellentéte a *termésátlagok* mutatója, ahol a vetítési alap, a termőterület legtöbbször hektárban adott). Jól ismert, hogy a legtöbb *demográfiai* mutatószámot (születés, halálozás, vándorlás) 1000 főre vetítve használják, s ezért *ezrelékben* adják meg (ez azzal függ össze, hogy a népességdinamika e mutatói jellemzően az adott népességszám egy-két százaléka körüli nagyságrendűek). A legtöbb, a népességszámhoz viszonyítással kapott fajlagost egy főre számítják. Bizonyos ellátottsági, infrastrukturális adatoknál gyakori a száz főre (száz háztartásra, lakásra) vetítés (ami azt is jelenti, hogy a kapott érték egyben azt mutatja, hogy a számba vettek hány százaléka él az adott ellátottsággal, pl. fürdőszobás lakással). Használati tárgyak esetében is gyakori a 100 főre vetítés (pl. személygépkocsi), sőt előfordulnak 1000 lakosra vetített mutatószámok is. Ezek makroszinten nagyon jól használhatók, máskor azonban furcsa eredményt adnak. A TV készülékek ezer lakosra jutó száma egy 250 fős aprófaluban mindenesetre érdekes fajlagos (ha minden lakosra jut egy készülék, akkor ez az adat itt 1000 s nem 250).

Indexek

Gyakran nagyon megkönnyíti az adatsorok értelmezését, megsokszorozza a következtetéseket, ha az eredeti adatokat az adatsor meghatározott értékeihez viszonyítjuk, *indexekké* alakítjuk. (Természetesen itt is figyelemmel kell lenni arra, hogy a viszonyításnak matematikailag is legyen értelme.)

Idősorok esetében a két leggyakoribban átalakítás:

- a *bázisindex*-idősor előállítása. Ekkor az adatsor minden értékét a kezdő időpont értékeihez viszonyítjuk (vagy elosztunk minden adatot az induló értékkel, vagy annak százalékában fejezzük ki, előbbi esetben a bázisidőponthoz rendelt érték 1, utóbbiban 100 lesz).
- a *láncindex*-idősorban minden értéket az előző időpontra vonatkozó adathoz viszonyítjuk (vagy elosztunk minden adatot a megelőzővel, vagy annak százalékában fejezzük ki)

A lánc-indexszé átalakítás sajátossága, hogy *megrövidíti* az idősort, hisz az induló időpont adatát nincs mihez viszonyítani. *A lánc-index idősor tehát nem 1-gyel vagy 100-zal kezdődik, hanem a kezdő időponthoz nincs hozzárendelt érték!*

Az idősorokhoz hasonló átalakítással juthatunk a *területi indexekhez.*

Nagyon gyakori a területi elemzésben az adatsor átlagához viszonyítás (lásd: egy főre jutó GDP az országos átlag százalékában). Ez az átalakítás az idősorok bázis-indexével mutat rokonságot. Területi lánc-index azonban nincs, hisz a területi adatsorokban nincs lineáris egymásutániság (az időbeli egymásutániságnak a térbeli szemléletben a *szomszédság* fogalma feleltethető meg). Az indexszé alakítás során eltűnik az adatsorok eredeti mértékegysége. Ez azzal a jótékony következménnyel jár, hogy a különböző mértékegységben mért adatsorok is összehasonlíthatóvá válnak. Ez fontos feltétele a komplex mutatók előállításának is (→ **5.4**).

Az indexszámítás kapcsán a fentiekben többször előkerült a *százalékos* adatforma. Ennek kapcsán érdemes felhívni a figyelmet a mértékkülönbségek vagy a változások százalékos illetve *százalékpontos* formában való kifejezésére. Utóbbi akkor kerül szóba, ha százalékos (arány) formában megadott adatok *különbségét* tekintjük jelzőszámnak. Ha például a városi népesség aránya 50%-ról 75%-ra nő, akkor a növekedés *25 százalékpontos* illetve *50 százalékos*, ugyanígy a munkanélküliségi ráta megduplázódása (azaz 100 százalékos növekedése) mondjuk 7%-ról 14%-ra, 7 százalékpontos növekedésként is interpretálható. Mindkét összevetés használható, de nem összekeverendő!

Megoszlási viszonyszámok

Abszolút adatok (például terület, népességszám, összes jövedelem, kereskedelmi forgalom, kukoricatermés) vizsgálatakor fontos alapinformációkat kaphatunk akkor, ha kiszámítjuk, hogy az egyes területegységekben (például a megyékben) mekkora hányada összpontosul az adott mutató (országos) volumenének. Ehhez az egyes területegységek százalékos részesedését kell kiszámítanunk. Az így módon kapott adatsort nevezzük (területi) megoszlási viszonyszám-adatsornak (a százalékszámítás bonyolult metódusát ismertnek tekintjük).

Mire jó ez az átalakítás? Néhány példa:

- Ha több évre kiszámítjuk egy adott jellemző megoszlási viszonyszámát, meghatározhatók a növekvő illetve csökkenő részesedésű térségek, a területi koncentrálódás illetve szétterülés folyamata.
- Míg az eredeti mértékegységben adott különböző jellemzők összehasonlítása a területegységek szerint nagyon nehézkes, ha megoszlási viszonyszámokká alakítjuk őket, összevethetők lesznek, eldönthető nagyságviszonyuk. (Azonnal eldönthető például így, hogy egy terület részesedése a népességből kisebb vagy nagyobb, mint a jövedelemből.)
- Megoszlási viszonyszámokkal is végezhetők műveletek. Ha például a népességrészesedést osztjuk a területaránnyal, akkor a népsűrűségre kapunk sajátos értéket. A „sajátos" jelző azt jelenti, hogy ekkor az országos népsűrűségi érték 1 lesz (mivel a 100/100 hányados adja), ennél kisebb értékek az átlagnál kisebb benépesültségű, nagyobbak sűrűbb térséget jeleznek (ha a hányadosokat 100-zal megszorozzuk, akkor az országos átlag 100 lesz, s ennek megfelelően módosulnak az egyedi adatok).
- Értékadatok (pl. jövedelem) esetében – ahogy arra már utaltunk – gyakori az, hogy különböző évek nominális, az áremelkedés, infláció által torzított adatai közvetlenül nem összehasonlíthatók. Ha azonban megoszlási viszonyszámokat számolunk belőlük, itt is azonnal eldönthető lesz, mely térség növelte jövedelemrészesedését, melyek a visszaszoruló, depressziós területek.
- Megoszlási viszonyszámokon alapul számos kiemelkedő fontosságú, gyakran használt területi egyenlőtlenségi mutató (→ **3.2.3**).

A „szumma"

A területi elemzésekben használt statisztikai összefüggésekben a leggyakrabban előforduló matematikai művelet a területi megfigyelési egységekre (legyen ezek száma n) vonatkozó számszerű adatok, jellemzők (jelölje ezeket x_i vagy y_i) *összeadása (szummázása).* Ennek a csak nagyon hosszú formában leírható műveletnek az egyszerű jelölésére használják a görög nagy Szigma betűt (Σ), így:

$$x_1 + x_2 + \ldots + x_i + \ldots x_n = \sum_{i=1}^{n} x_i$$

szavakkal: „i egyenlő 1-től n-ig szumma x_i," ami n darab szám (az adott indikátornak a területi egységekre vonatkozó értékei) összeadását jelenti az elsőtől az utolsóig.

Az ilyen összefüggésben a szumma jel *után* szerepelnek az összeadandó értékek, a szumma jel alatti és feletti *indexjelölés* pedig azt mutatja meg, hogy melyik és hány adatot kell összeadnunk. A legtöbb esetben az összes (területi) megfigyelési egység adatát össze kell adni, ebben az esetben leegyszerűsödik a jelölés, a *szumma jel indexei elhagyhatók*, ekkor az alábbi kifejezések egymással azonos jelentésűek:

$$\sum_{i=1}^{n} x_i = \sum_{i} x_i = \sum x_i$$

A szumma jel alsó és felső indexeit mindenképp jelölni kell, ha az adatok összeadása csak a megfigyelési egységek egy részére terjed ki.

Nem csak az esetenként különböző értékű egyedi változók, hanem azok különböző matematikai összefüggéseinek (összegeik, szorzataik vagy más függvényeik) összeadására is szükség lehet. Ilyenkor az alábbi jelölések és összefüggések érvényesek:

Ha az adott változó értéke minden területegységben ugyanazt a konstans értéket (*a*) veszi fel, akkor az alábbi összefüggéssel van dolgunk (mivel a konstansok minden megfigyelési egységben ugyanazon értékűek, ezekhez alsó index természetesen nem tartozik):

$$\sum a = na$$

Ha az adott változót (x_i) egy tetszőleges állandóval, konstanssal (legyen ennek jele *b*) megszorozzuk, majd ezt követően adjuk össze az adatokat, akkor a konstans kiemelhető:

$$\sum bx_i = b\sum x_i$$

Ha két változó összegét szummázzuk minden megfigyelési egységre, akkor két utat is választhatunk. Vagy először minden megfigyelési egységre összeadjuk a két adatot (x_i-t és y_i-t) és ezt követően szummázunk, vagy előbb külön-külön összegezzük az egyes adatsorokat majd ezután adjuk össze a két teljes összeget. A végeredmény ugyanaz:

$$\sum (x_i + y_i) = \sum x_i + \sum y_i$$

Hasonló logika alapján egyszerűsödik az alábbi összefüggés is, ha *b* konstans:

$$\sum (x_i - b) = \sum x_i - nb$$

Nagyon fontos felhívni a figyelmet arra, hogy nem áll fenn ugyanilyen jellegű összefüggés az adatsorok szorzatösszegeinél. Ekkor:

$$\sum x_i y_i \neq \sum x_i \sum y_i$$

azaz szorzatok összege nem (feltétlenül) egyenlő az összegek szorzatával. Ez alábbi egyszerű számpéldával könnyen belátható.

Mintapélda az összegek szorzatának és a szorzatok összegének különbségére

Területek(i)	x_i	y_i	x_iy_i
1	0	1	0
2	1	0	0
3	0	2	0
4	2	0	0
5	0	3	0
6	3	0	0
7	0	4	0
8	4	0	0
9	0	5	0
Σ	**10**	**15**	**0**

ahol $\sum x_i y_i = 0 \neq \sum x_i \sum y_i = 10 * 15 = 150$

A szummázások esetében nagyon fontos a zárójelek precíz használata, ezt példázzák az alábbi összefüggések:

$$\sum (x_i - b) = \sum x_i - nb \neq \sum x_i - b$$

vagy

$$\sum x_i^2 \neq \left(\sum x_i\right)^2$$

ami azt mutatja, hogy számok négyzeteinek összege nem egyenlő számok összegeinek négyzetével. Egyszerű algebrai, majd számpéldával, két elemre bemutatva:

$$x_1^2 + x_2^2 \neq (x_1 + x_2)^2 = x_1^2 + 2x_1x_2 + x_2^2$$

$$1^2 + 5^2 = 26 \neq (1+5)^2 = 36$$

Ezekre az összefüggésekre tekintettel nagy figyelemmel kell számítani olyan gyakori mutatószámokat, mint az adatsorok szórása vagy a korrelációs együtthatók!

Összeadni nem csak egy-egy adatsor (vektor) elemeit lehet, hanem egy mátrixét is. Ez a gyakorlatban kétféleképp történhet, először soronként haladva adjuk össze az elemeket, majd a sorösszegeket is összeadjuk vagy ugyanígy teszünk, de oszloponként haladva. Ha a képletekben két szummát látunk közvetlenül egymás mögött, akkor ott egy mátrix elemei kerülnek összeadásra:

$_{\mathbf{n}}\mathbf{A}_{\mathbf{m}}$ elemei összeadásának jelölése: $\sum_i \sum_j x_{ij} = \sum_j \sum_i x_{ij}$

A Σ (n diszkrét adat összeadása) matematikai kiterjesztése *folytonos* változókra az integrálszámítás területére vezet.

A szummához hasonló jelölési logikával egyszerűsíthető le n szám szorzata (a szorzás jele itt:*). A jelölésben a nagy görög Π (pí, p) betűt használjuk, így:

$$x_1 * x_2 * \ldots * x_i * \ldots * x_n = \prod_{i=1}^{n} x_i$$

2.2.4. Azonosság, egyenlőség, egyenlőtlenség

A matematikában gyakran kapcsolunk össze kifejezéseket relációs jelekkel. Az azonosságok, az egyenletek és az egyenlőtlenségek olyan matematikai relációk, amelyek két oldalán ismeretlenekből (változók), konstansokból (állandók) és műveletekből álló összefüggések szerepelnek.

Azonosságnak azt a relációt nevezzük, amelynek két oldalán szereplő kifejezések az ismeretlenek (ez alábbi esetben a és b) tetszőleges értéke mellett, azaz mindig egyenlők:

$$(a+b)^2 = a^2 + 2ab + b^2$$

Azonosságok esetében a két kifejezés összekapcsolására az ≡ (azonosság) jelet is használják.

Az egyenletekben a két, egyenlőségjellel (=) összekapcsolt kifejezés közötti egyenlőség csak a változók (ismeretlenek) meghatározott értékei esetében áll fenn. Az egyenletek „megoldása” épp ezen értékek megkeresését jelenti. (Vannak olyan egyenletek, amelyeknek általában, vagy a számok valamely körére vonatkozóan nincs megoldása.)

Az *egyenlőtlenségek* az összekapcsolt kifejezések közötti nagyságviszonyt fejezik ki, a változók (ismeretlenek) meghatározott értékei (értelmezési tartománya) mellett. A-val és B-vel jelölve az összekapcsolt kifejezéseket az alábbi egyenlőtlenségi relációk lehetségesek:

A mindig *kisebb, mint* B (ami analóg azzal, hogy B mindig nagyobb, mint A):

A < B és B > A

A mindig kisebb vagy egyenlő, mint B (ami analóg azzal, hogy B mindig nagyobb vagy egyenlő, mint A):

A ≤ B és B ≥ A

2.2.5 Függvények

Legáltalánosabb matematikai értelemben két halmaz elemeinek *egymáshoz rendelését* (vagy más szavakkal X halmaznak Y halmazra való leképezését) *függvénynek* nevezzük, ha van olyan *utasítás* (szabály, képlet, táblázat, ábra stb.), amely X halmaz minden egyes *x* elemének megfelelteti Y halmaz

egy *f(x)*-szel jelölt elemét. Az X halmazt *értelmezési tartománynak*, Y-nak az összes *f(x)*-ből álló részhalmazát *értékkészletnek* nevezzük.

A függvényeket – például az utasítás, a hozzárendelési szabály jellemzői, tulajdonságai szerint - sokfajta csoportba, típusba sorolhatjuk. A függvény fogalma – tág értelemben – nem jelenti azt, hogy X halmaz valamely értékéhez Y halmaznak egy és csak egy értéke rendelhető, bár ez a leggyakoribb eset. *Többértékű függvénnyel* azonban a területi elemzésben is találkozhatunk, ilyen például a földrajzi és az időtávolság összefüggését leíró kapcsolatot, két azonos légvonal-távolságú pontpár a hálózati kiépítettség különbözősége okán nagyon erltérő időtávolságú lehet.

A függvénytan alapvető fogalmaival a matematikai *analízis* foglalkozik. Ennek része a *differenciál- és integrálszámítás*, amely, egyebek mellett a függvények tulajdonságainak (alakjának, szélsőértékeinek, maximumainak és minimumainak stb.) meghatározásához alapvető jelentőségű matematikai összefüggéseket kínál. A szélsőérték-számítás, optimumkeresés sok matematikai-statisztikai eljárásban is alapeszköz, csak egyet említve, ezen alapszik szinte az egész regresszió-analízis is, ahol a mért és becsült adatok eltérését számszerűsítő *legkisebb négyzetek módszere* szélsőérték-számításra vezet → **4.4**.

Lineáris kapcsolat

Ha *x* minden egységnyi növekedéséhez, *x* értékétől függetlenül *f(x)*-nek *b értékű* (b állandó, konstans*)* növekedése tartozik, akkor a két halmaz között lineáris függvénykapcsolat van, a függvényt derékszögű koordinátarendszerben ábrázolva *egyenest* kapunk. A közismert képlet:

$$y = a + bx$$

ahol *x* a *független*, *y* a *függő* változó, b az ún. *meredekség*, a pedig (ami az *x*=0 értékhez tartozó *y* érték) az ún. *tengelymetszet* (az egyenes b magasságban metszi az y tengelyt). Ha *b* együttható pozitív, akkor x növekedésével y is növekszik, ha negatív, akkor csökken, ha értéke 0, akkor az x tengellyel párhuzamos, *a* magasságban lévő egyenes a függvénykép.

Ha a független változók halmaza nem egy-, hanem kétdimenziós, akkor lineáris összefüggés esetében a függő változó értékei egy síkot képeznek. Ilyen összefüggést ír le az alábbi képlet:

$$z = a + bx + cy$$

3 független változó esetén a függvény képe egy hasáb, n független változó esetében pedig egy (vizuálisan, grafikusan nem interpretálható) *n-dimenziós „sík”*. Az n vektorból ($\mathbf{x_1,\ldots x_n}$) és n skalárból ($a_1 \ldots a_n$) képzett $\sum_{i=1}^{n} a_i x_i$ összefüggést (vektort) a vektorok *lineáris kombinációjának* nevezzük (ilyen összefüggés van például a faktoranalízis módszerében az alapváltozók és a faktorok között → **5.5**).

A logaritmus

Az $a^x = b$ exponenciális egyenlet megoldásakor olyan kitevőt keresünk, amelyre a alapot felemelve b-t kapunk. Az egyenletnek a > 0, a ≠ 1 és b > 0 esetén van egy és csak egy megoldása. A b pozitív szám a alapú logaritmusának nevezzük azt a kitevőt, amelyre a-t emelve b-t kapunk. Jelölése: $\log_a b$ (a alapú logaritmus b).

A logaritmus definíciójából következik, hogy $\log_a 1=0$ *és* $\log_a a =1$.

A logaritmus alapja tetszőleges 1-től különböző pozitív szám lehet. A gyakorlati (statisztikai, elemző) munkában leggyakrabban a 10-es alapú logaritmust használjuk, s (index nélkül) log-gal jelöljük. Mivel többfajta előnyös matematikai tulajdonsága van, gyakran használják az ún *természetes alapú logaritmust* (ln), ahol az alap e=2,71828, ami az $(1+1/n)^n$ függvény határértéke.

Az ún. logaritmikus azonosságok közül (emlékeztetőül): számok szorzatának logaritmusa egyenlő a számok logaritmusainak összegével, azaz:

$$\log(\prod_i x_i) = \sum_i \log x_i$$

Számos jelenség növekedése exponenciális jellegű, egy-egy időszak alatt a volumenek hatványozódnak. E függvények logaritmusát véve a függvénygörbe képe egyenest ad (ezzel függ össze az, hogy gyakran használják az ún. szemi-logaritmikus koordinátarendszert, amelynek függőleges (y) tengelyén a beosztás nem lineáris, hanem logaritmikus).

Sok elemzés igazolja, hogy a lélekszám, a településnagyság fontos fejlettségi tényező, ezért gyakorta szerepel összetettebb elemzésekben is, így a *regresszióelemzések* → **4.4** magyarázó változóiként. Ott gyakran feltűnik az, hogy nem a lélekszámot, hanem annak logaritmusát szerepeltetik a számításban. E transzformáció matematikai következménye, hogy a gyakorta nagyon széles intervallumon eloszló településnagyság logaritmizálva jobban közelíti a matematikai-statisztikában elvárt *normális* eloszlást, s így a számítások matematikai szempontból sem kérdőjelezhetők meg. A lélekszám (általában minden abszolút volumen) logaritmusának konkrét jelentése is van, a települések *nagyságrendjét* tükrözi, így tartalmilag is megfelelően értelmezhető a hatása, szerepe: 10-es alapú logaritmusa esetén log 1 = 0, log 10= 1, log 100= 2, log 1000 = 3.

2.3 Adattípusok, mérési skálák

A statisztikában a megismerni kívánt, megfigyelt egységek halmazát *sokaságnak* nevezik. A sokaság jellemzésére, részekre bontására alkalmas vizsgálati szempontok az *ismérvek.* A megfigyelt sokaság egyes tagjaihoz az ismérvek szerint különböző *ismérvértékek* tartoznak.

A területi elemzés mindig legalább két ismérvvel dolgozik. Implicit módon – már a megfigyelési egységek kiválasztásakor – megjelenik a *területi* ismérv, majd ehhez kapcsolódva az ismérvek más típusai: *időbeli, mennyiségi, minőségi* jellemzők. Ezeket v*áltozónak*, esetenként csak a *számszerű* ismérveknek nevezik.

A legkülönbözőbb megfigyelési egységekre vonatkozó mérések, adatok, információk jól csoportosíthatók aszerint, hogy milyen összehasonlításra alkalmasak. Ezt a csoportosítást tartalmazza a *mérési (vagy adat) skálák* vagy *mérési szintek* rendszere (összefoglalóan lásd a *2.2. táblázatban*).

Legyen adott két megfigyelési egység, A és B, s ezekre vonatkozó két mért vagy számbavett jellemző, ismérv: X_a és X_b.

Vannak olyan jellemzők, amelyek esetében csak az dönthető el, hogy A és B e jellemző szerint *megegyezik egymással vagy különbözik egymástól.* Az ilyen jellemzőket *nominális* skálán mért adatoknak vagy nominális ismérveknek nevezzük. E skála esetében két vagy több értéke, kimenete lehet a változóknak.

Ennél már finomabb összevetésre van módunk A és B között, amikor nemcsak a megegyezés vagy a különbözőség állapítható meg, hanem a *nagyságviszony*, a sorrend is, azaz, hogy A vagy B jellemzője a nagyobb. Ez az ún. *ordinális (sorrendi) skála.*

Még finomabb összevetésre van mód, amikor a megegyezés - meg nem egyezés, valamint a sorrendiség megállapításán túl az is kideríthető az adatokból, hogy mennyivel nagyobb az egyik érték a másiknál (értelmezhető az adatok *különbsége* is). Az ilyen jellemzőket *intervallum-skálán* mért adatoknak nevezzük.

A legteljesebb összehasonlításra az ún. *arány-skálán* mért adatok esetében van mód. Ekkor az összes fenti reláción túlmenően értelmezhető az is, hogy egyik adat hányszorosa a másiknak (megadható a két adat *aránya*).A mérési skálák sajátos hierarchiát alkotnak, amelynek csúcsán az arányskála áll. Minden skálára igaz az, hogy az előző skálák tulajdonságaival rendelkezik.

Annak ismerete, hogy valamely jellemző milyen skálán mért, azért fontos, mert bizonyos statisztikai jellemzők és matematikai-statisztikai eljárások csak bizonyos adatskálákra számíthatók, illetve alkalmazhatók. Pusztán statisztikai alapon merészség azt állítani, hogy a magyar fociválogatott 63-szor rosszabb, mint a listavezető brazil (ha a két csapat az 1. illetve a 63. helyen áll a ranglistán), egy 0 átlagú adatsort pedig nem is lehet az átlag százalékában megadni.

Mérési skálák	**Tulajdonság**	**Értelmezhető relációk**	**Sajátosságok**	**Jellemző példák**
Nominális	Megkülönböztetés	$X_a = X_b$ vagy $X_a \neq X_b$	Nem számszerű	Név, születési hely, nem
Ordinális	Megkülönböztetés, sorrend	$X_a = X_b$ vagy $X_a \neq X_b$ és $X_a \geq X_b$ vagy $X_a < X_b$	Nehezen mérhető, csak sorrendbe állítható	Sorrendek, (katonai) rangok
Intervallum	Megkülönböztetés, sorrend, különbség	$X_a = X_b$ vagy $X_a \neq X_b$ és $X_a \geq X_b$ vagy $X_a < X_b$ Értelmezhető $X_a - X_b$	Pozitív és negatív értékek	Vándorlási különbözet, átlaghőmérséklet
Arány	Megkülönböztetés, sorrend, különbség, arány	$X_a = X_b$ vagy $X_a \neq X_b$ és $X_a \geq X_b$ vagy $X_a < X_b$ értelmezhető $X_a - X_b$ valamint X_a/X_b	Van elméleti minimum, azonos előjelű	Népességszám, jövedelem

2.2. táblázat *A mérési skálák rendszere*

Az adatskálák közötti átmenetek közül talán a leggyakoribb az intervallum- vagy arányskálán mért jellemzők *sorrendi (ordinális) adatskálára történő átalakítása*. (Ez az adat-transzformáció *nem az adattábla sorainak, azaz az egyes területegységekhez tartozó értékek sorrendjének felcserélését jelenti*, – elemi szinten gyakori hiba és súlyos félreértés – hanem egy új adatsor létrehozását, ahol az eredeti adatsor legnagyobb értékéhez 1, a következőhöz 2 (vagy fordított rangsorolás esetén a legkisebbhez 1, a következőhöz 2 és így tovább) számérték kerül, új adatként. Előfordulhat az, hogy az átalakításra kerülő adatsorban megegyező értékek szerepelnek, az ezekhez rendelt rangszámok esetében előnyös fenntartani azt az összefüggést, hogy a rangszámok összege ne változzék. Ez adategyezés esetében úgy biztosítható, hogy páros számú egyezés esetében a rangszámok átlagát rendeljük a megegyező adatokhoz, míg páratlan számú adat egyezésekor a középső rangszámot. Ha például a 4. és 5. helyen álló érték azonos, akkor a két rangszám egyaránt 4,5, ha a 8., 9. és 10. helyen álló érték egyezik meg, akkor mindhárom megfigyelési egységhez a 9 rangszámot rendeljük. A statisztika tehát nem ismer „holtversenyt", egy adatsor két azonos, legnagyobb értékéhez a rangsorolásban nem 1 – 1 érték – „két aranyérem" – rendelődik, hanem 1,5 – 1,5, ezáltal biztosítható, hogy a rangszám összeg ne változzék, ugyanis: $1,5 + 1,5 = 3 \neq 1+1 = 2$.

Érdemes még felhívni a figyelmet arra, hogy amennyiben minden alapadat különbözik, akkor nincs jelentősége annak, hogy a sorrendi átalakítást *csökkenő vagy növekvő* rendben végezzük el. Azonos, a fenti szabályok szerint átalakítandó értékeket is tartalmazó adatsorok esetében azonban a kétfajta rangsorolás szerint új adatsornak ugyan az átlaga azonos, de a *szórása* különbözhet.

A rangszámok és az eredeti értékek közötti összefüggés önmagában is vizsgálatra érdemes. Ezzel foglalkozik a nagyon sok társadalmi jelenségben, közte a mi szakmai körünkben elsősorban a településrendszerek vizsgálatakor tesztelhető, ún. *rang-nagyság szabály* →**4.4.7**).

2.4 Kódolás

A kód valamely információ egyértelmű visszaadására alkalmas, megállapodás szerinti jel vagy szimbólum. A hétköznapi élet is tele van kódolt információkkal. (Az egész távközlés és informatika működésének alapfeltétele a nagyon különböző jellegű információk átvihető formájú jelekké alakítása.) A kódok előállítása a kódolás, a kódokból az eredeti információk előállítása a dekódolás.

A kódoknak a területi elemzésben leggyakrabban az alábbi két jelentése van:

- mint a területi megfigyelési egységek *azonosítói*
- mint *transzformált, átalakított* (kódolt) *adatok*

A területi azonosítók, bár részei az adattábláknak, elemzési szempontból mégis más minőséget képviselnek, mint a jelzőszámok, adatok, indikátorok. A területi azonosítókkal műveleteknek csak nagyon korlátozott köre végezhető el, lényegében két reláció megállapítására alkalmasak: az eredeti információk azonosságának vagy különbségének megállapítására, valamint (jól kiválasztott azonosítók esetében) a megfigyelési egységek rendezésére, csoportosítására. Más számtani műveleteket azonban nem szabad végezni velük (átlaguk például értelmetlen). Mindebből következően ugyancsak viszonylag csekély (ha határozottabban fogalmazunk: semmi) értelme új tudományos eredmények reményében a település- vagy térségazonosító számjeleket korreláltatni a jövedelmekkel. Mindez nem kérdőjelezi meg azt, hogy a területi azonosítók a (mennyiségi) elemzések nélkülözhetetlen statisztikai eszközei.

A településkód és a KSH területi számjelrendszere (Kiss János Péter)

A hazai területi statisztikában használt KSH-településkód a települések egyedi meghatározására szolgáló sorszám jellegű, tartalom nélküli (a rendszer létrehozásakor, 1981-ben véletlenszerűen kiosztott) egyedi azonosító, amely a település alakulásától megszűnéséig változatlan, és a település megszűnése után sem használható fel más település azonosítására. A településkód eredeti formájában egy 5 számjegyből álló karakterlánc (szöveg formátum), ám az utolsó jegyet, az ún. CDV-számot (amely csak a számítógépes ellenőrzést szolgálja), gyakran elhagyják, az első jegy pedig 0 is lehet. Így a településazonosító törzsszámok – numerikus formátumban – gyakran 201 és 3500 közötti három- vagy négyjegyű számokként jelennek meg. (A településkód szerinti sorban legelső település, a Fejér megyei Ráckeresztúr esetében például az eredeti 5 jegyű településkód a 02015 karakterlánc, amelynek numerikus megfelelője 2015, a CDV-szám elhagyásával képzett, leggyakrabban használt verzióban pedig 201. Budapest esetében ugyanezek rendre 13578 (karakterlánc), 13578 (szám), 1357 (szám).) A kapott településsoros adatállományok használatakor erre oda kell figyelni, és szükség esetén a megfelelő konverziót elvégezni (a legegyértelműbb, így legcélszerűbb a CDV-szám elhagyásával képzett numerikus kód használata, további példáink ezeket mutatják).

Fontos megjegyezni, hogy – a rendszer egységét némiképp megbontva – Budapest kerületei és a főváros egészében is rendelkeznek saját településkóddal (Budapest 1357-es kódja mellett pl. a IV. ker. kódja 546). Ez kedvező a Budapesttel foglalkozók számára, ugyanakkor az adatbázisok használata során mindig meg kell győződni arról, hogy a vizsgált adatbázisban milyen módon szerepel Budapest, és szükség esetén a 23 kerület adatait össze kell vonni, illetve a duplán (kerületenként és összesen) is szereplő adatok egyikét törölni. Az adatok teljeskörűsége érdekében néhány fiktív területegységet is „létrehoztak”, azaz kóddal láttak el a rendszerben. Ezek – Budapest kerületre nem bontható, a megyék településre nem bontható, az ország területre nem bontható adatai, illetve az országhatáron kívüli tevékenység – kétjegyű numerikus kódokkal rendelkeznek. Ha a hozzájuk tartozó adatérték nem nulla, akkor megfelelő kezelésükre, esetleges kiszűrésükre figyelni kell (pl. az országos összesítéseknél). A közigazgatási változások – településösszevonások, szétválások – miatt a számozás nem folyamatos (pl. a 2005. jan. 1-jei állapot szerint „legutolsó”, 3145. település, a 2002-ben önállósodott Somogy megyei Zákányfalu törzsszáma 3440).

A településkód a Központi Statisztikai Hivatal 1981-ben bevezetett, a különböző területi egységek és a hozzájuk tartozó települések egyértelmű azonosítását célzó területi számjelrendszerének része. A legújabb, 2004. jan. 1-jétől alkalmazott számjelrendszerben – amely összesen 57 számjegyből áll – az alábbi besorolásokat rögzítették minden településre, illetve Budapest kerületeire vonatkozóan:

- az 5 jegyű településazonosító törzsszám (1-5. pozíció)
- a területi jelzőszám (6-11. pozíció), amelynek részei:
    - a megyekód (2 számjegy);
    - a főváros kerületeinek kódja (2 számjegy);
    - a település jogállásának kódja (1 számjegy);
    - megyeszékhely-kód (1 számjegy);
- kistérségkód (4 számjegy, 12-15. pozíció), illetve hogy a település kistérségi központ-e (0 – nem, 1 – igen): 1 számjegy, 16. pozíció;
- régiókód (1 számjegy, 17. pozíció);
- agglomeráció-kód (4 számjegy, 18-21. pozíció; az első jegy az agglomeráció típusára utal, a második-harmadik számjegy azonosítja az adott agglomerációt, míg a 4. jegy az adott agglomeráción belül különbözteti meg a központot, az esetleges társközpontokat és a többi települést);
- körjegyzőség-kód (összesen 6 számjegy, 22-27. pozíció; közülük az első jegy fejezi ki, hogy körjegyzőség székhelye, tagja, vagy önálló jegyzőséggel rendelkezik-e a település; a további 5 jegy az érintett jegyzőség központjának 5 jegyű településkódja);
- a mezőgazdasági tájkörzet kódja (3 számjegy, 28-30. pozíció);
- a szőlő-termőtáj és a borvidék kódja (3 számjegy, 31-33. pozíció);
- a vállalkozási övezet kódja (2 számjegy, 34-35. pozíció)
- az idegenforgalmi régió kódja (2 számjegy, 36-37. pozíció)
- az üdülőkörzet kódja (2 számjegy, 38-39. pozíció);
- a világörökségi helyszín kódja (2 számjegy, 40-41. pozíció)
- a nemzeti park kódja (2 számjegy, 42-43. pozíció)
- kulturális látványosság, esemény kódja (2 számjegy, 44-45. pozíció)
- az aktív turizmus kódja (2 számjegy, 46-47. pozíció)
- a gyógyturizmus kódja (3 számjegy, 48-50. pozíció)
- ún. NUTS-kód az Európai Unió számára történő adatszolgáltatások biztosításához (7 számjegy, 51-57. pozíció).

A területi számjelrendszer elemeinek segítségével végezhető a legcélszerűbben a települési és területi adatok csoportosítása, elemzése és tematikus térképeken való ábrázolása. Egyrészt ugyanis a T-STAR mellett a KSH valamennyi egyéb elektronikus adatközlése, de egyre gyakrabban más adatszolgáltatók is erre támaszkodnak, ezt használják, másrészt numerikus jellege miatt jóval gyorsabb és pontosabb, mint pl. a nevek segítségével történő azonosítás. Egyszerűsége és szabvány „mivolta" miatt pedig megkönnyíti az adatok cseréjét: segítségével szinte bármilyen forrásból származó adatsort „házilag" is könnyen hozzá lehet kapcsolni meglevő adatainkhoz. (A települések számjelei papíralapú KSH-kiadványban és elektronikus formában is – külön, vagy a T-STAR-hoz kapcsolva – egyszerűen beszerezhetők a KSH-tól, újabban a KSH honlapjáról is letölthetők.) Mindezek miatt az adatok számítógépes feldolgozása és térképezése során is érdemes a hivatalos település-, kistérség-, megye- és régiókódokat szerepeltetni z adatsorok elején.

A rendszert 6 település (a főváros, egy budapesti kerület, egy megyeszékhely és három, különböző megyékhez tartozó község) példáján érzékeltetjük:

Helységnév	Budapest	Zugló (Bp. 14. ker.)	Nyíregy-háza	Szany	Cered	Vámos-család
Településazonosító törzsszám	1357 8	1633 7	1720 6	0853 6	0366 5	3105 1
Területi jelzőszám	0100 1 1	0114 9 9	1500 2 1	0800 4 2	1200 4 2	1800 4 2
Kistérségkód	3101 1	9999 9	4508 1	3801 0	4205 0	4806 0
Régiókód	1	1	6	3	5	3
Agglomeráció kódja	0001	0001	3051	9999	9999	9999
Körjegyzőségi kód	1 13578	9 99999	1 17206	2 08536	2 03665	3 21537
Mezőgazdasági tájkörzet kódja	621	621	224	514	112	524
Szőlő-termőtáj és borvidék kódja	999	999	999	999	999	999
Vállalkozási övezet kódja	99	99	99	99	07	99
Idegenforgalmi régió kódja	01	01	03	09	02	09
Üdülőkörzeti kód	01	01	99	99	25	99
Világörökségi helyszín kódja	01	99	99	99	99	99
Nemzeti park kódja	99	99	99	99	99	06
Kulturális látványosság, esemény kódja	11	99	31	99	99	99
Aktív turizmus kódja	01	99	10	99	99	99
Gyógyturizmus kódja	202	999	204	999	999	999
NUTS-kód	HU10101	HU99999	HU32308	HU22101	HU31305	HU22206

A t*elepülésazonosító törzsszám* a már ismertetett ötjegyű településkód. A gyakorlatban – mint azt már említettük – elégséges és célszerűbb az *első négy értékes számjegy* használata (Budapest esetében 1357, Cerednél 366 és így tovább).

A *területi jelzőszám* a magasabb közigazgatási egységekhez való tartózás hatjegyű kódja. Elemei közül önállóan is gyakran használatos a *megyé*t azonosító első két jegy, a megyekód (a 01-es sorszám Budapesté, majd a 02-es Baranyától a 20-as Zaláig nevük ábécérendjében következnek a megyék, Jász-Nagykun-Szolnokot kivéve, amelynek megyekódja a megye 1990 előtti megnevezése (Szolnok) alapján 16, a 15-ös Szabolcs-Szatmár-Bereg és a 17-es Tolna között). A következő két érték a *fővárosi kerületek* azonosítására szolgál (01-től 23-ig, a többi településre vonatkozóan – így Budapest egészére is – ez az érték 00). A jelzőszám ötödik tagja *a települési jogállást* jelzi (1: főváros, 2: megyei jogú város, 3: város, 4: község), a hatodik érték pedig a *megyeszékhely-kód* (1: megyeszékhely, 2 nem megyeszékhely). Ennek megfelelően Nyíregyháza kódja azt jelzi, hogy a település Szabolcs-Szatmár-Bereg megyében található (15), megyei jogú város (2) és megyeszékhely (1). Szany Győr-Moson-Sopron megyei (08) község (4), Zugló pedig Budapest (01) 14. kerülete, amelynek esetében a települési jogállás és a megyeszékhely-kód nem értelmezhető (Erre itt a 9-es számjegy utal, a számjelrendszer más elemeiben a 99, 999, 9999-es kódok jelzik az adott szempontból nem érintett településeket).

A *kistérségkód* a hazai térségi elemzés ma talán legfontosabb területegységeibe, az ország 168 kistérségébe sorolja be a településeket. Budapest itt a 3101-es sorszámmal az első, Baranya megye komlói kistérsége a 3201-es, mohácsi kistérsége a 3202-es stb. (a sort Zala megye zalaszentgróti kistérsége zárja 5006-os számmal. A Függelék kistérségi adattábláiban megtalálható valamennyi kistérség kódja és neve.) Példáink közül Cered (4205) a salgótarjáni, Szany (3801) a csornai kistérséghez tartozik. A kód ötödik jegye 1 (kistérségközpont) és 0 (nem központi település) lehet (a felsoroltak közül pl. Nyíregyháza kistérségközpont, Cered és Szany pedig nem).

A *régiókód* a 7 tervezési-statisztikai régióhoz tartozást azonosítja: Nyíregyháza Észak-Alföld (6), Zugló a Közép-Magyarországi Régió (1) része. A *NUTS-kód* az EU statisztikai rendszerét követi. 1-2. értéke az országazonosító (HU) a 3. érték a NUTS 1 szintre utal (ez 1 – Közép-Magyarország –, 2 – Dunántúl –, vagy 3 – Alföld és Észak – lehet). A 4. pozíció a NUTS 2 (régiók) szintje az adott NUTS 1 egységen belül 0 és 3 közötti értékekkel, a következő a megyék kódja (a régión belül), míg az utolsó érték a kistérségi kód (a megyén belül). Azaz ez is hierarchikus kód. A fentiek értelmében tehát Vámoscsalád NUTS-kódjából kiderül, hogy a község Magyarországon (HU), ezen belül a Dunántúlon (2), annak is a Nyugat-Magyarországi régiójában (2), azon belül pedig Vas megye (2) sárvári kistérségében (06) található. (A települést is azonosító teljes, 11 jegyű kódot a NUTS-rendszerben a településazonosító törzsszám 1-4. pozíciójával kiegészített NUTS-kód adja.) A *vállalkozásiövezet-kód* a területfejlesztés sajátos, preferált területegységeit azonosítja (a rendeletileg meghatározott övezetek száma 1999 óta 11). A táblázatban szereplő települések közül csupán Cered érintett ilyenben: Salgótarján–Bátonyterenye Térsége Vállalkozási Övezet (07).

A számjelrendszer egyes elemei a települések sajátos igazgatási funkcióit illetve besorolását jellemzik. Így a *körjegyzőségkód* a helyi igazgatási szerep szerint osztályoz, s jelzi, hogy például Budapest és Nyíregyháza olyan települések, ahol önálló, csak az adott városra kiterjedő jegyzőség működik (1). Hozzájuk hasonlóan Szanynak és Cerednek is saját maga a jegyzőségi székhelye, ám esetükben más településeket is ellátó körjegyzőség működik (2), míg Vámoscsalád körjegyzőség nem-székhely települése (3), amelynek illetékes körjegyzője a 2153-as településkódú Uraiújfalun dolgozik. Az *agglomerációkód* a KSH által hivatalosan lehatárolt népességtömörülésekhez való tartozást jelzi. Zugló például a budapesti agglomeráción (0) belül, Budapest (00) része, amely az agglomeráció központi települése (1), míg Nyíregyháza egy nagyvárosi településegyüttes (3), mégpedig a nyíregyházi (05) része, sőt központja (1). Ugyanakkor egyetlen, példáink között szereplő község sem tartozik agglomerálódó térséghez (9999).

A *mezőgazdasági tájkörzet* kódja egy háromszintű hierarchikus rendszerbe sorolja be a településeket a mezőgazdaság szempontjából releváns adottságaik szerint (makro-, mezo- és mikrokörzetek). Így Nyíregyháza ebből a szempontból a Nyírség makrokörzethez (2), ezen belül a Nyírség mezokörzethez (2) – a másik mezokörzet itt a Felső-Tiszavidék (1) – és az Észak-Nyírség mikrokörzethez (4) sorolt. Szany esetében Kisalföld (5), Kisalföld (1) és Rábaköz (4) a három hierarchia-szint. *Nemzeti parkhoz* tartozó területe példáink közül csak Vámoscsaládnak van (06, Fertő-Hanság Nemzeti Park), egyik település sem tartozik ugyanakkor 22 *szőlőtermő és borvidékünk* valamelyikébe (999).

A területi számjel többi eleme az idegenforgalmi karakterről ad információt. Valamelyik *idegenforgalmi régiónkba* minden települést besoroltak – a főváros például a Budapest-Közép-dunavidékibe (01) tartozik –, *üdülőkörzetbe* viszont a fővároson kívül csak Cered tartozik példáink közül (mégpedig a 25-ös kódú „Cserhát és környéke" elnevezésűbe). *Világörökségi helyszíne* a hat településből csak a fővárosnak van, mégpedig „Budapest Duna-parti látképe, a Budai Várnegyed, az Andrássy út és történelmi környezete" (01), érdemi *kulturális* vonzereje viszont Nyíregyházának is („Jelentősebb vonzerejű kulturális látványosság konferenciahelyszínnel", 31-es kód). Az *aktív* (=lovas vagy vízi) *turizmust* szolgáló létesítményei, illetve *gyógyturisztikai* adottságai ugyancsak Budapestnek és Nyíregyházának vannak a felsoroltak közül[3].

A KSH a települések egyedi kódjait a közigazgatási térfelosztás, a településosztályozás, illetve az egyes idegenforgalmi adottságok változásakor ennek megfelelően módosítja. (2005-ben például újabb 13 település kapott városi rangot – a Dunántúlról Őriszentpéter, Martonvásár, Kadarkút, a fővárosi agglomerációból Üllő, Pilis és Ócsa, az Alföldről Nyírtelek, Nyírlugos, Kisköre, Abádszalók, Csorvás, Sándorfalva valamint a zempléni Pálháza – ezek esetében megváltozik *a település jogállásának kódja*.) Amennyiben alapvető módosításra kerül a hazai területi közigazgatás egész rendszerében, akkor ugyancsak változni fognak a területi kódok, bár az osztályozás alaplogikáját, főbb elemeit ez vélhetően nem fogja felborítani.

A KSH számjelrendszerén kívüli legismertebb területi azonosító a *postai irányítószámok* rendszere (Vigyázat: nem minden településnek van egyedi postai irányítószáma!). Területi, települési azonosítókat, kódok maga a kutató is kreálhat. A települési lélekszám is kódolható, nagyságcsoportokba sorolhatóak a települések, kóddal lehet jellemezni a sajátos fekvési pozíciókat (pl. határmentiség) vagy akár különböző térségtípusokat is.. A kódrendszer folyamatosan követi a bekövetkező közigazgatási változásokat, amelyek a *Helységnévtárakban* követhetők nyomon.

Adatátalakítás kódolással

Az *adatátalakítás* során használt kódolás legjellemzőbb esete a *nominális változók számszerűsítése*, egyértelmű szabály szerinti, kódolt formára alakítása (ez a kódolás az empirikus szociológiában a leggyakoribb).

Speciális eset a két kimenetű minőségi változók átalakítása bináris kódokkal 0, 1 értékekből álló adatsorrá. Az ilyen jellemzők összetettebb matematikai-statisztikai modellekbe is beépíthetőek. Ennek az adja meg az alapját, hogy az ilyen adatsor átlaga nem fiktív szám csupán, hanem jól értelmezhető tartalommal bír (az 1 értékű egységek arányát méri). Ugyanez nem teljesül a többértékű nominális változók kódolásakor.

[3] Ahogy a fenti leírás is mutatja a KSH számjelrendszere messze nem valamifajta semleges statisztikai jeltár, hanem egy sokelemű *„hivatalos" településosztályozási rendszer,* amely sokfajta területi vizsgálatban kiindulópont lehet.

Ilyen bináris jellegűek a területi elemzésben (például a regresszió számításban nagyon jól használható) ún. *dummy* változók is. A magyarországi településekhez ugyanis hozzárendelhető 7 különböző regionális dummy: minden település abban a regionális vagy (régió-) dummyban 1 értéket kap, amely régióban található, a másik hatban pedig 0-t. A legutóbbi táblázatban, a területazonosítók kapcsán szereplő települések közül Szany például a Nyugat-Dunántúl változójában (dummyjában) kaphat 1-t, a többiben pedig 0-t, Cered értéke ellenben az Észak-Magyarország dummyban 1, a többiben 0. Azt már csak ellenőrző kérdésként tesszük fel: mi a helyzet Nyíregyházával?

Több kimenetű minőségi változók esetében már bonyolultabb a kódolás. Ha a nominális tartalomhoz nem vagy csak nagyon erőltetetten lehet legalább sorrendiséget, kisebb-nagyobb, fontos-lényegtelen, támogatott-elutasított stb. relációt rendelni, akkor ugyan elvégezhető a számszerűsítés, de az új adatsorral való bonyolultabb számításoknak már nagy a kockázata. A főváros – megyeszékhely – város – község „értékeket" tartalmazó nominális változóhoz ugyan logikusnak látszik az 1, 2, 3, 4 kódok hozzárendelése, de mindebből már nem következik az, hogy e településtípusok között bármilyen értelemben is 1 lenne a különbség, amit a számszerűsítés sugall. A véleménykutatások esetében nagyon gyakori *elégedettség- vagy attitűd-vizsgálatok* esetében[4] hasonló a helyzet: a nagyon elégedett – elégedett – elégedetlen – nagyon elégedetlen skálához rendelt 4, 3, 2, 1 számok esetén. Itt ráadásul még az a faktor is felbukkan, hogy lehetséges (sőt biztos) az, hogy vannak olyan interjúalanyok, akinek az elégedetlensége körülbelül azonos viszonyt jelent a vizsgált jelenséghez, mint mások elégedettsége (mindez tovább fokozódik a bővebb 0-10 vagy 0-100 közötti skálákon mozgó értékeléseknél).

A kódolás azonban nem pusztán nominális jellemzők számszerűsítését jelentheti. Hanem eszköze lehet az *információ-tömörítésnek* is: széles intervallumon eloszló változók esetében már a számítások során összevont csoportokat képezhetünk, majd ezekkel végezhetjük a számítást. (Ez a logika mintegy fordítottja a fentinek, s leggyakrabban akkor kerül szóba, ha az adott számszerű jellemző „puha", nem közvetlenül számba vett, s így egyedi értékeinek megbízhatósága kérdéses, ami a csoportokba rendezéssel csökkenthető.)

2.5 Matematikai-statisztikai alapok

A *statisztika* sokjelentésű fogalom. Ha azt halljuk, „statisztikát készít", akkor számokkal bűvészkedő hivatalnok juthat eszünkbe. A „statisztikai adatgyűjtés" már nagy szervezettséget igénylő, tervszerű adat-előállítás képét hívja elő. Az adatgyűjtés fogalma kapcsán érdemes különbséget tenni kísérletes adatgyűjtés és a „valódi" statisztikai munka között, előbbit bár sokszor számokat eredményez, nem hívjuk statisztikának. Az adatfeldolgozás bizonyos módszerei önmagukban is „statisztikák".

A mennyiségi módszertant használó területi vizsgálatok jellegadó információi a különböző forrásokból származó *statisztikai* adatok. A társadalom- és gazdaságstatisztikai adatokra támaszkodó elemzéseknek a statisztikával szembeni közvélekedés jellegzetes negatív megnyilvánulásaival is szembe kell néznie (*J. Best* amerikai szociológus szerint az emberek négyféle módon tekintenek a statisztikai adatokra: *félelemmel telve, naív értetlenséggel, cinikusan vagy szélsőségesen kritikusan.*) A társadalomkutató esetében az utóbbi (a szélsőséges jelző nélkül) átvehető a mindennapi vélekedésből. Arra is érdemes ügyelni, hogy maga az elemzés, adatfeldolgozás ne rontsa tovább a statisztika megtépázottnak tűnő tekintélyét, hanem segítse az adatokban rejlő mondandó felszínre hozatalát.

A *matematikai-statisztika* a többfázisú statisztikai munka adatfeldolgozási, elemzési fázisához kacsolódó, alapvetően elméleti, a matematikához erősen kötődő tudományterület, ami azt vizsgálja, hogy az összegyűjtött adatokból milyen módon, milyen feltételekkel vonhatunk le következtetéseket. A matematikai-statisztikai összefüggések *tömegesen* előforduló jelenségekre vonatkoznak, belőlük közvetlen következtetések az egyes egyedi esetekre nem erednek.

---

[4] A legismertebb, leggyakrabban alkalmazott 5 illetve 7 fokozatú ordinális (a „negatív", elutasító pólustól a semleges ítélet fokozatain át a „pozitív" elfogadó véleményig húzódó) skálát *Rensis Likert* (1903-1981), a társadalomtudomány számos ágában működő amerikai kutató dolgozta ki és publikálta 1932-ben, őróla kapta nevét az ún. *Likert-skála*.

2.5.1 Valószínűségszámítási háttér

A valószínűségszámítás a matematikának viszonylag fiatal, néhány száz éves ága (kialakulásában a szerencsejátékoknak és a biztosítási kockázatok elemzésének volt kitüntetett szerepe), ami mára az egész tudományos elemzési kultúrát, mind a természet-, mind a társadalomtudományok terén átszőtte. A területi elemzések matematikai-statisztikai alapelemei és bonyolultabb eszközei mögött is mindenütt ott található a valószínűségszámítási háttér.

Szemléletileg arról a felismerésről van szó, hogy a különböző jelenségek (a valószínűségszámítás alapkategóriája szerint, *események, kísérletek)* kimenete, bekövetkezte, felbukkanása, nem merev ok-okozati, ún. determinisztikus kapcsolatok folyománya, amikor az okból egyértelműen következik az egyértelmű okozat. Sokkal gyakoribb, hogy a kapcsolat valószínűségi jellegű (sztochasztikus), ami azonban – bizonyos feltételek ismeretében – nem hagy teljes bizonytalanságban, megmondható (előre jelezhető, kiszámítható), hogy egyik vagy másik kimenet, esemény gyakori vagy ritka, nagy vagy kicsiny a valószínűsége.

Nincs itt mód a valószínűségszámítás legelemibb fogalmainak részletezésére sem, de példaként két nagyon egyszerű gondolatsort azért vehetünk.

Először egy történeti példa a determinisztikus és a sztochasztikus kapcsolatra. Ha 1966-ban bárhol bementem a közértbe s a 3,60-as kenyérből 4 kilót vettem, akkor 14 Ft 40 fillért fizettem (determisztikus kapcsolat), ha volt kenyér (sztochasztikus kapcsolat a boltba menetel és a kenyérvétel között). Ma a kenyérvásárlás épp ellentétesen determinisztikus, illetve sztochasztikus: kenyér (általában) mindig van, csak az ára nem rögzített, s így a fizetendő árra csak valószínű becslés adható.

Az események különböző kimeneteit, azok valószínűsége alapján a (0,1) intervallumon jellemezzük. A *biztos* esemény (például, hogy a feldobott – nem űrrakétával fellőtt - 100 forintos visszahull a földre) valószínűsége 1, a *lehetetlen* esemény (hogy a földre hullott 100 forintos érmén 50 forintos felirat látható) valószínűsége 0. Egyaránt 0,5 (vagy másként 50%) a valószínűsége a „fejnek” és az „írásnak”. A lehetséges kimenetek valószínűségének összege 1, ebből következően az adott példa harmadik lehetséges kimenete (megáll az élén az érme) *0 valószínűségű (de nem lehetetlen*, ami azt jelenti, hogy sok ezer érmefeldobás során bizony bekövetkezhet, ha nagyon ritkán is. Ugyanilyen, 0 valószínűségű, bár nem lehetetlen esemény az is, amikor valaki az előadások látogatása nélkül jelesre kollokvál statisztikából.

Egy esemény bekövetkeztének valószínűsége, bizonyos (ritka) esetekben tisztán matematikai (leggyakrabban kombinatorikai alapon) kiszámítható, megadható, máskor csak kísérletek hosszú sorának ismétlésével, s a kimenetek összegzésével becsülhető meg.

A valószínűségszámítás kapcsolódik egy másik rendkívül felértékelődött, önálló tudományterülethez, az *információelmélethez* is (lásd pl. *Theil, H. 1970*). Ennek ugyanis egyik alapfogalma az, hogy minél kisebb egy esemény valószínűsége, annál nagyobb az információtartalma (ami azt jelenti, hogy annál meglepőbb, ha előfordul). Ezt az alapösszefüggést használják ki a területi egyenlőtlenségek információelméleti mérőszámai → **3.2.3**.

A valószínűségszámítás és a matematikai-statisztika, illetve a statisztikai módszertan különbségét leegyszerűsítve úgy összegezhetjük, hogy bár utóbbiak az előbbi elméleti bázisára épülnek, de jóval gyakorlatiasabbak, ami alapvetően azt jelenti, hogy nem elméleti függvényekkel, hanem véges *mintákkal* (az elméleti eseményteret közelítő kimenet-sorokkal) dolgoznak. Ez a reláció emeli a matematikai-statisztika középponti fogalmává a különböző megbízhatósági vagy *szignifikancia* vizsgálatokat, amelyek arra valók, hogy segítségükkel ellenőrizhessük, hogy az adott minta tekinthető-e az elméleti függvény (eloszlás) megfelelő realizációjának (reprezentásának). Itt kell ismételten utalnunk egyik korábbi megjegyzésünkre, miszerint a kvantitatív elemzésekben és a matematikai-statisztikában ma sincs nyugvóponton az a vita, hogy minden adatsort mintának kell-e tekinteni vagy elemezhető anélkül is, hogy a szignifikancia-probléma fel se merüljön. A gyakorlati területi statisztikai elemzésekben ez utóbbi álláspont az uralkodó, a megbízhatóság, szignifikancia kérdése azonban még ezen feltevés mellett is felbukkan abban a követelményben, hogy *megbízható összefüggések jellemzően csak nagy elemszámú (hosszú) adatsorból számíthatók.*

Az alábbi táblázatban összefoglaljuk a statisztikai elemzésekben leggyakrabban előforduló fogalmaknak, elnevezéseknek megfelelő, azokkal *rokon jelentésű* valószínűségszámítási alapfogalmakat, ezeket néha – kissé pongyolán - felváltva is használják.

Statisztika	*Valószínűségszámítás*
Adatsor, jellemző, változó	Valószínűségi változó
Relatív gyakoriság	Valószínűség
Átlag, középérték	Várható érték
Szórás	Szórás
Gyakorisági hisztogram	Sűrűség-függvény
Eloszlás	Eloszlás-függvény
Minta	Minta, reprezentáció
Idősor	Sztochasztikus folyamat

2.3. táblázat *A statisztika és a valószínűségszámítás rokon fogalmai*

2.5.2 Adatsorok statisztikai jellemzői, jellegadó értékei

Középértékek

N szám (adat) számtani átlaga az a szám, amelyet az eredeti számok helyébe helyettesítve a számok összege változatlan, azaz a *számtani átlag* (x „felülvonás" jelöléssel):

$$\bar{x} = \frac{\sum x_i}{n}, \text{ ami ekvivalens azzal, hogy } n\bar{x} = \sum x_i$$

N fajlagos adat (y_i) *súlyozott számtani átlaga*, ahol a súlyok jelölése f_i (súlyként a fajlagos mutató nevezőjében szereplő adat – az egy főre jutó jövedelem esetében a népesség - megfelelő területi értékei illetve megoszlási viszonyszámai, azaz a megfelelő területegységeknek a teljes népességből való százalékos részesedései egyaránt használhatók):

$$\bar{y} = \frac{\sum y_i f_i}{\sum f_i}$$

A súlyozott formula tulajdonképp semmi más, mint összevont formája annak, hogy minden egyedi adatot (még, ha az területileg aggregált formában, például területegységekre összevontan adott is) figyelembe veszünk a számításnál. Ha súlyozatlan átlagot számítunk n területegység aggregált fajlagosából, akkor tulajdonképp egy olyan n elemű minta átlagát kapjuk, amely az egyes területegységek átlagos értékeiből áll. (Ugyanez a helyzet a súlyozatlan illetve súlyozott szórás esetében → **2.5.3**). A területi egyenlőtlenség-vizsgálatokban jellemzően súlyozott átlagokat, illetve szórást használnak, számos matematikai-statisztikai módszer ellenben a területi adatokat csak súlyozatlan formában képes kezelni.

N szám *mértani átlaga* az a szám, amelyet a számok helyébe helyettesítve azok szorzata változatlan (x „csillag" jelöléssel):

$$x^* = \sqrt[n]{\prod x_i} \text{ ami ekvivalens azzal, hogy } \left(x^*\right)^n = \prod x_i$$

A számtani és mértani átlag között határozott nagyságviszony van, a számtani átlag mindig nagyobb vagy egyenlő, mint a mértani. Két tagra az alábbi levezetés bizonyítja a tételt:

$$\frac{a+b}{2} \geq \sqrt{ab}$$

$$a + b \geq 2\sqrt{ab}$$

$$(a+b)^2 \geq 4ab$$

$$a^2 + 2ab + b^2 \geq 4ab$$

$$a^2 - 2ab + b^2 \geq 0$$

$$(a-b)^2 \geq 0$$

$$q.e.d.$$

Az adatsorok (valószínűségi változók) helyzeti középértékei közé tartozik a *medián,* ami az az érték, aminél kisebb és nagyobb adatok száma egyenlő. A helyzeti középértékeket extrém adatokat tartalmazó adatsorok esetében előnyös használni.

A medián tulajdonképp egyik sajátos esete az adatsor *kvantiliseinek*, amelyek olyan értékek, amelyek az adatsor nagyság szerinti sorrendbe rendezett tagjait azonos elemszámú részekre tagolják. A medián a „felező" érték, gyakran használják a kvartiliseket („negyedelő" pontok), a kvintiliseket (ötödölők), a deciliseket („tizedelő") illetve a percentiliseket („századolók"). Ezeknél beszélünk 1, 2,...k-adik (k a kvartilisnél negyedik, a decilisnél tizedik, percentilisnél a századik hányadot jelenti) kvantilisről. Ezek a jellemzők különösen gyakran fordulnak elő a jövedelem-egyenlőtlenségek vizsgálatakor („felső tized").

A jövedelem-eloszlások esetében (amelyek jellemzően lognormálisak) a medián/átlag arány maga is egyenlőtlenségi mutatóként fogható fel. Minél kisebb e hányados értéke, annál nagyobbak a jövedelemegyenlőtlenségek.

Sem az átlag, sem a medián nem feltétlenül egyezik meg az adatsor valamely tagjával (ezért mondhatjuk joggal, hogy nincs „átlagmagyar" csak „magyar átlag"). A „felső kvartilis" például az az érték, aminél az adatsor tagjainak egy negyede nagyobb úgy, hogy maga az adott érték nem is biztos, hogy szerepel az adatsorban (folytonos változók esetében a kvantilisekre csak becslések adhatók). Ezzel szemben, az ugyancsak a középértékek közé sorolt a *módus* (vagy módusz, a „divatos" érték), *az adatsornak az az értéke, ami legtöbbször* fordul elő. Ha több ilyen érték is található az adatsorban, akkor azt többmódusúnak (többcsúcsú) nevezzük.

Terjedelem, szórás

Az adatsorok jellegadó értékei közé tartozik a *maximális* és a *minimális* érték (akár egyik, akár másik többször is előfordulhat az adatsorban). A maximális és a minimális érték különbsége az adatsor *terjedelme* (range-arány). Használható a differenciáltság mérésére a maximális és a minimális érték különbségét (a szóródás terjedelme)

A matematika-statisztika egyik legfontosabb fogalma a *szórás*, ami az adatsorok egyes értékeinek az átlagtól való négyzetes eltérésének az átlaga, képletben:

$$s = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n}}$$

Fajlagos adatok esetében legtöbbször a szórás súlyozott formáját számítjuk:

$$s = \sqrt{\frac{\sum_i (y_i - \bar{y})^2 f_i}{\sum_i f_i}}$$

Az adatsorok szórása 0 és +∞ közötti értékeket vehet fel, mértékegysége megegyezik az eredeti adatsoréval. (A szórásra a területi egyenlőtlenségi mutatók → **3.2.2** kapcsán még visszatérünk.)

Amennyiben egy *minta* alapján *becsüljük* egy feltételezett eloszlás szórását, akkor – valószínűségszámítási megfontolások alapján – a kifejezés nevezőjében n helyett n-1 áll (ezt korrigált szórásnak hívják), ami arra utal, hogy a mintákból számított értékek általában kissé alábecslik a teljes sokaság szórását. Számos statisztikai programcsomag ezen összefüggések alapján számol, s ezért kis elemszámú adatsorok esetén a két érték jelentősen el is térhet. Ha azonban pl. 150 kistérség valamely adatából számítunk szórást, a 150-nel vagy 149-cel való osztás az értékes jegyekben már nem okoz különbséget. Még inkább igaz ez a súlyozott formulára. Ez is példa arra, minél nagyobb a minta (az adatsor hossza), annál megbízhatóbbak a belőle számítható értékek.

Ha a szórást az átlag százalékában fejezzük ki (csak intervallum-skálán mért, nem 0 átlagú adatok esetében tehetjük ezt meg), kapjuk a *relatív szórást*:

$$v = \frac{s}{\bar{x}} * 100$$

Mivel az előzőekben bemutatott 3 középérték közül a módus minden mérési skála esetén értelmezhető, a medián az ordinális skálától kezdődően, az átlag csak az intervallum- és arányskálán, szórást is csak ilyen mérési szintű

adatokból számíthatunk. (Azaz nem átlagolhatók sem a nemek, sem a katonai rangok, de megadható, hogy egy csoportban férfi vagy nő van több, s milyen rangnál kezdődik a hadsereg felső tizede.)

A szórás az adatsorok értékeinek differenciáltságát, tagoltságát jellemzi, s fontos mutatója a területi egyenlőtlenségek mérésének is → **3.2.1** Az egyenlőtlenségvizsgálatokban nagy hátránya a szórás-mutatónak, hogy felülről nem korlátos (bármilyen nagy lehet, ezért értéke alapján nem dönthető el, hogy „nagy" vagy "kicsi" az adatok differenciáltsága).

Standardizálás

Az átlagot és a szórást használja az adatsorok egyik alapvető, *dimenziótlanító eljárása* az ún. standardizálás. X_i adatsorból az alábbi összefüggés alapján állítható elő Z_i standardizált változó:

$$z_i = \frac{x_i - \overline{x}}{s_x}$$

ahol $\overline{x}$ az adatsor átlaga s_x az adatsor szórása. A standardizált változó (z_i) átlaga 0, szórása 1.

Ahhoz, hogy igazoljuk, hogy a standardizált változó átlag valóban 0, szórása pedig 1, fel kell írni z_i átlagának illetve szórásának képleteit, abba behelyettesítendő a z_i és x_i közötti fenti kapcsolat, majd algebrai átalakításokkal eljuthatunk az eredményhez.

Eloszlások

A különböző adatsorok (vagy általánosabban a valószínűségi változók) többféleképp csoportosíthatók matematikai tulajdonságaik, fő jellemzőik (például átlaguk, szórásuk) alapján.

Két nagy csoport mindenképp megkülönböztetendő:

- a *diszkrét eloszlások* értelmezési tartománya diszkrét, egyedi értékekből áll (nem folytonos). Ilyen például a családok eloszlása a *családtagok száma* szerint egy országban (1, 2, 3 …n tagú családok száma). Magától értetődően diszkrét minden nominális vagy ordinális skálán mért jellemző is. A diszkért eloszlásból számított átlagok tulajdonképp csak fiktív adatok (pl. ilyen szám az, hogy egy országban az átlagos családnagyság 2, 79, ennek azonban azonnal értelmes jelentés adunk akkor, ha azt mondjuk, hogy itt 100 család átlagosan 279 főből áll).
- a *folytonos eloszlású* változó egy adott intevallumon belül tetszőleges értéket felvehet. Ilyen mutató például a *jövedelem*, ami (elvileg) bármilyen nem negatív értéket felvehet (ha a jövedelem nélküli, csak adóssággal rendelkező egyedektől eltekintünk), de hasonló a hőmérséklet, a csapadék, a távolságmértékek, a tengerszint feletti magasság stb. Az adott intervallum „kitöltése" csak a mérés pontosságától és a megfigyelések számától függ.

A folytonos eloszlások fogalmához kapcsolódik a *sűrűség- illetve az eloszlásfüggvény*. Előbbi a relatív gyakoriságokat, utóbbi azok kumulált értékét ábrázolja az értelmezési tartományon. Az empirikus elemzésben a sűrűségfüggvénynek felel meg az adott minta értékeinek oszlopdiagramon (→ **2.9.2**) történő ábrázolása, az ún. *gyakorisági hisztogram*[5]. A leggyakrabban használt eloszlások közül néhány külön kiemelést érdemel. [6]

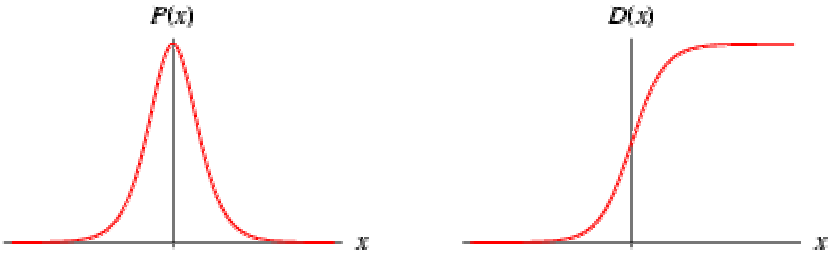
[5] Véges mintával reprezentált folytonos eloszlások sűrűségfüggvényének (amely „lépcsős") kisimítására használható az ún. *kernel-módszer*.

[6] Az interneten 118 eloszlás matematikai leírását tartalmazza a http://mathworld.wolfram.com/topics/StatisticalDistributions.html weboldal.

Normális eloszlás

Soktényezős hatások eredményeként létrejövő kimeneteleket leíró szimmetrikus eloszlás, („harang alakú” eloszlás, Gauss-görbe). Statisztikus jellemzője az, hogy a normális eloszlású adatok kétharmada az átlagtól egy szórásnyi távolságon belülre esik, s két szórásnyi intervallumon belül van az adatok 98 százaléka *(2.2. ábra).*

A valószínűségszámításban illetve a matematikai-statisztikában ez az egyik leggyakrabban előforduló eloszlás. A diszkrét binomiális eloszlásból származtatható. Több összetett matematikai-statisztikai összefüggés normális eloszlást feltételez.

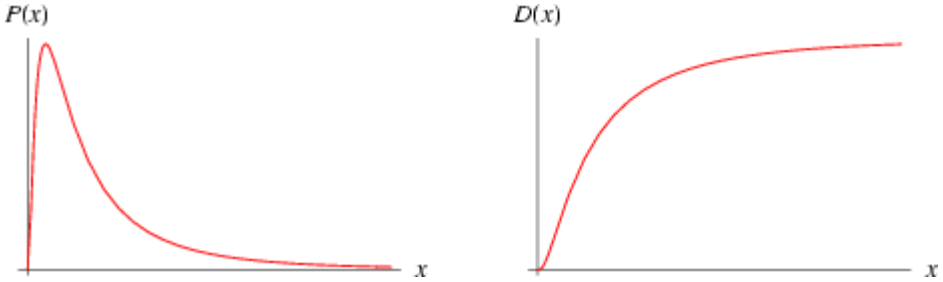


2.2.. ábra A normális eloszlás sűrűség- és eloszlásfüggvényének képe

Bár az empirikus vizsgálatok területi adatsoraira sokkal inkább az a jellemző, hogy eloszlásuk nem normális (sűrűségfüggvényük nem harang alakú, nem Gauss-görbe), több (pl. k darab) ilyen adatsor együttes eloszlása már közelíti a normális eloszlást a k-dimenziós térben (a területegységek többsége a k-dimenziós ponthalmaz közepe körül tömörül).

Lognormális eloszlás

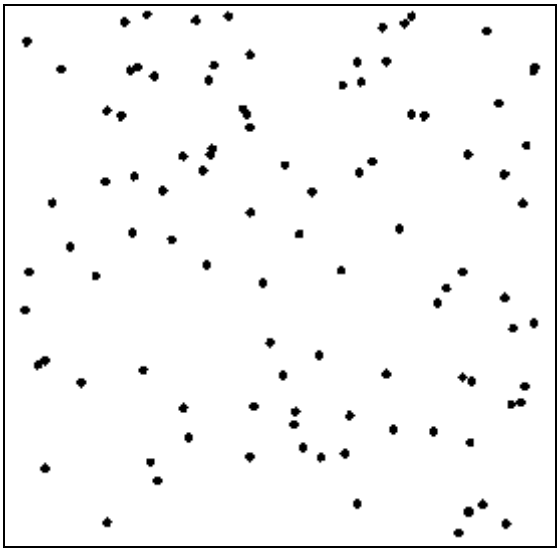
Ha egy adatsor (valószínűségi változó) logaritmusának eloszlása normális, akkor az eloszlást *lognormális* eloszlásnak nevezzük *(2.3. ábra).* Ez az eloszlás a természeti és társadalmi jelenségek esetén nagyon gyakori, nem szimmetrikus (balra ferde) eloszlás, ahol a kis értékek gyakoriak, a nagyok ritkák, de messzire elnyúlik az eloszlás. Ilyen a természetföldrajzból ismert szemcseeloszlás vagy a jövedelmek eloszlása: sok „szegény” (kis és közepes jövedelmű) és jóval kevesebb magas jövedelmű embert találunk mindenütt. A jövedelemeloszlás jellege mindenütt ilyen, csak a gazdag-szegény polarizáció mértéke (a jövedelmek szórása) különbözik koronként és országonként, a gazdag országok jellemzően kiegyenlítettebbek, mint a szegények. A lognormalitással ekvivalens az a kijelentés, hogy *az adatok nagyságrendje normális eloszlású.*



2.3.. ábra A lognormális eloszlás sűrűség- és eloszlásfüggvényének képe

További eloszlások

A számos további eloszlás közül a *Poisson eloszlás* (*Simeon D. Poisson*, francia matematikus, 1781-1840) fordul elő leggyakrabban a területi elemzések során. Ez írja le például *a térben (síkon) véletlenszerűen eloszló pontok* (*2.4. ábra*) konfigurációját (annak valószínűségét, hogy egy adott térrészbe 0, 1 2…n pont esik). Térbeli mintavételek esetében az ilyen konfiguráció az egyik megfelelő választás. Az ábrán is jól látszik, hogy a „véletlenszerűség” úgy jelentkezi, hogy egyaránt találunk egymáshoz nagyon közeli, sűrűsödő pontokat s viszonylag üres térrészeket. Ehhez az eloszláshoz viszonyítva számszerűen is jellemezhető a ponteloszlások konfigurációja (koncentráltsága, véletlenszerűsége, szabályos, geometrikus jellege) a legközelebbi szomszéd analízis módszerével (*Nemes Nagy J. 1998*)



2.4. ábra A véletlen kétdimenziós ponteloszlás (Poisson eloszlás) képe (*Rosenberg 2001* nyomán)

2.6 Térbeli középértékek, súlypont

Egy n pontból álló síkbeli pontrendszer súlypontjának koordinátái, ha a pontok helyzete koordinátarendszerben (térképen) adott, és minden ponthoz egy-egy „súly" (tömeg) tartozik, a pontok koordinátáinak súlyozott számtani átlagaként számíthatók:

$$x = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}; \quad y = \frac{\sum_{i=1}^{n} f_i y_i}{\sum_{i=1}^{n} f_i}$$

A fenti összefüggésben x és y a súlypont két koordinátáját, x_i és y_i az alappontok koordinátáit, f_i pedig az alappontokhoz tartozó súlyokat jelöli. Ha a súlyok azonosak, akkor a pontrendszer *geometriai* súlypontját adja meg a számítás. Súlypont számításához tehát az alappontok helykoordinátáira és az alappontokhoz rendelt súlyra (tömegre) van szükség.

A módszer a területi elemzések klasszikus eszköztárába tartozik. Alkalmazásának gyökerei a 20. század első felének *szociálfizikai* elemzési irányzataiig vezethetők vissza *(Stewart, J. Q. – Warntz, W. 1958)*. Az 1920-as években a Mengyelejev Centrográfiai Laboratóriumban, Leningrádban kiemelt kutatási témaként szerepelt. A súlypontmodellnek megfeleltethető bármely területegység, ország népességének területi megoszlása; a pontoknak a települések, a súlyoknak a népességszámok tekintendők. A *népességi* súlypont kiszámításával a területi megoszlás egyik középértékét kapjuk meg, a súlypontnak az alappontoktól mért euklideszi távolságai négyzetének súlyozott összege minimális. Magyarországon a súlypont-modell használatára épült *Bene L. – Tekse K. 1966* nagyívű empirikus elemzése a hazai népesség területi eloszlásának történeti változásairól.

A számításokban súlyként nemcsak a népesség, hanem bármely más "tömegmérték" is szerepeltethető. Ilyen lehet például számos gazdasági indikátor helyi volumene: foglalkoztatottak száma, termelési érték, jövedelem. Ezekkel számolva különböző *gazdasági* súlypontokat, a gazdasági tevékenység térbeli középpontjait kapjuk eredményül. A gazdaságon kívüli társadalmi jelenségek vizsgálatára is alkalmas a módszer: a politikai földrajzi elemzésekben a különböző pártokra adott szavazatok lehetnek a súlyok (ezekkel számolva egyfajta „politikai" súlypontokhoz juthatunk), de vizsgálhatók a módszerrel akár a társadalmi devianciák (bűncselekmények, öngyilkosság, balesetek) is.

Történetileg vizsgálva az egyes súlypontokat, képet kapunk a térbeli elmozdulások jellegzetes irányairól. A módszerrel látványosan igazolható például a népesség történeti távú kelet-nyugati mozgása, a benépesülési folyamat az Amerikai Egyesült Államokban. Az USA statisztikai hivatala minden népszámlálás kapcsán közli az ország népességi súlypontjának helyét is, s történeti statisztikai források alapján immár több mint 200 éves időtávon kísérhető nyomon a népességi súlypont mozgása.

Különböző súlypontok egymáshoz viszonyított távolsága és iránya szintetikus ismereteket ad fontos társadalmi térbeli folyamatok térszerkezetéről:

- a geometriai és a népességi súlypont viszonya a *népsűrűség* földrajzi differenciáltságát szintetizálja,
- a geometriai és a gazdasági súlypont viszonya a gazdaság (a termelés) területi *koncentrációjáról* nyújt információt,
- a népesség és a jövedelem súlypontjának viszonya a *jövedelemegyenlőtlenségek*, a gazdasági fejlettség differenciáltságát jellemzi.

Bár a súlypontelemzés kétségtelenül jól használható a térbeli arányeltolódások történeti elemzésében, bizonyos folyamatokat azonban elvileg sem képes kimutatni. Noha lényeges térszerkezeti változások következhetnek be egy területen, mégsem mozdul el a súlypont akkor, ha a változás (a súly-növekedés vagy -csökkenés) a súlypont körül szimmetrikusan zajlik. Ezek között az esetek között az alappontoknak a súlyponttól mért átlagos, súlyozott távolsága - az ún. *standard távolság* - tesz különbséget. Minél kisebb ennek értéke, annál jellemzőbb az, hogy az adott jelenség a súlypont körül koncentrálódik.

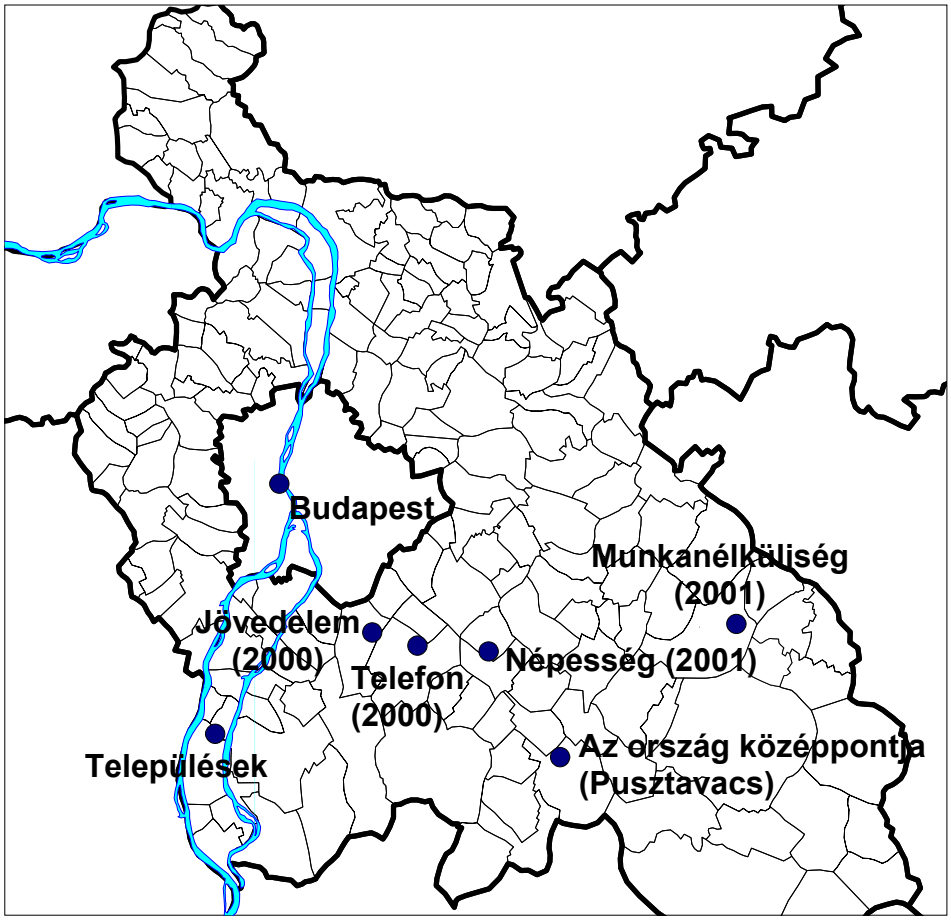
A standard távolság (D) számítása:

$$D = \sqrt{\sum_{i=1}^{n} f_i((x_i - x)^2 + (y_i - y)^2)} \Big/ \sum_{i=1}^{n} f_i$$

ahol x_i és y_i az alappontok koordinátáit, x és y a megfelelő súlypont koordinátáit, f_i pedig az egyes alappontokhoz tartozó megfelelő súlyokat (népességszám, munkanélküliek, telefonvonalak száma) jelöli.

Érdemes még említést tennünk egy térség, ország ún. (népességi vagy gazdasági) *medián-pontjáról* is. Ezt úgy kapjuk meg, hogy a térképeken megkeressük azokat az egyeneseket, amelyek K-Ny illetve É-D irányban „elfelezik” a népességet és a két egyenes (hosszúsági illetve szélességi kör) metszéspontja adja ki a mediánpontot. Ez a pont sokkal kevésbé érzékeny a népesség vagy bármely más vizsgált jelenség térbeli arányeltolódására, hisz egyáltalán nem mozdul el akkor, ha például a népesség eloszlásában a térség egy-egy felén belül történik elmozdulás. (Ha Tolna megye összes lakosa Zalába települne át, a népességi súlypont markánsan elmozdulna nyugati irányban, a mediánpont azonban változatlan maradna.)

Magyarország népességi és gazdasági súlypontjai az egész huszadik században *jellemzően a fővárostól délkeleti irányban* helyezkedtek el (tendenciájában egyre közelebb kerülve a fővároshoz, elsősorban annak népesség- és gazdaságkoncentráló ereje hatására). Ez a helyzet az ezredfordulón is. A 2000-es népességi súlypont *Vasad* község közelébe esik (Budapest „elméleti középpontjától”, a Lánchíd budai hídfőjénél lévő 0 kilométerkőtől légvonalban 33 km-re, DK-re). Az országos jövedelmi súlypont már nagyon megközelítette a fővárost, s *Alsónémedi* község központjához van legközelebb, 22,2 km-re a főváros középpontjától. (Településmarketing ötletként interpretálva: míg az ország földrajzi “közepe” Pusztavacs, addig a “jövedelemi közép” ma(!): Alsónémedi). Miként a jövedelemi súlypont, úgy a telefonvonalak súlypontja is főváros és a népességi súlypont közé esik Budapesttől 26,2 km-re (*Ócsa* közelébe). A munkanélküliségi súlypont az összes korábbitól messze elkülönülve, a fővárostól keleti irányban, attól 58,6 km-re adódott 2001-ben, *Tápiószele* közelében, még Pest megyében (további részletek: *Nemes Nagy J. 2003*).



Súlypont	*Súly*	*Év*	*K-Ny koordináta* (km)	*É-D koordináta* (km)
Felsőfokú végzettség	Felsőfokú végzettségűek elvégzett osztályszáma	2001	-8,26	-8,91
Települések	Minden település=1	2001	-7,58	-31,73
Jövedelem	Adóköteles jövedelem	2000	11,33	-19,06
Telefonellátottság	Telefon fővonalak	2000	16,74	-20,64
Iskolázottság	7-x évesek elvégzett osztályszám	2001	21,48	-21,39
Népesség	Lakónépesség	2001	25,03	-21,48
Földrajzi középpont	-	2001	34,50	-35,41
Munkanélküliek	Regisztrált munkanélküliek	2001	60,41	-16,41

2.5. ábra *Súlypontok a 20. század végén Magyarországon (Budapest origójú koordinátarendszerben)*

2.7. A területi mozgóátlag[7] (Dusek Tamás)

A gazdasági-társadalmi jelenségek térségi különbségeinek vizsgálatát rendszerint a közigazgatási határok keretein belül végzik. A másik lehetséges módszer az elemzési célokra kialakított egyéni lehatárolás keretében vizsgálódni. A harmadik lehetőség a területi mozgóátlag alkalmazása, amely módszer a lehatárolási probléma és az aggregációs információveszteség egyfajta kezelése mellett a

[7] A fejezet a módszernek csak a lényegét ismerteti, részleteiben lásd *Dusek T. 2004, 9. fejezetben* vagy a szerzőnek a módszert bemutató publikációjában (*Dusek T. 2001*).

térszemlélet formálására is alkalmas. A módszer szemléleti rokonságot mutat, a pont-mintavételekből történő területi szintű becslés, a térbeli interpoláció más módszereivel.

A közigazgatási határvonalakkal készült tematikus térképek szükségszerűen egyetlen homogén tömbként jelenítik meg az egyes területegységeket. Mindez nem okoz zavart, ha tudatosítjuk, hogy valójában a megfigyelési egységeken belül nem találkoznánk ilyen egyöntetű homogenitással, a határsávok többnyire inkább hasonlítanak egymásra, mint a területegységek távolabbi, belső térségei. „A természet nem szereti az ugrásokat" térbeli értelemben sem, a szokásos ábrázolás azonban elterelheti erről a figyelmet. Ennek bizonyítéka és egyben távoli következménye a területegységeket, mint homogén egységeket kezelő modellek gyakori használata és népszerűsége. Az ilyen modellek mellett egyszerűségük és könnyű kezelhetőségük szól, a térbeliség tartalmi jellemzői miatt azonban csak az elemzések első, meglehetősen általános lépcsőfokának célszerű tartani őket.

2.7.1 A területi mozgóátlag számítása és típusai

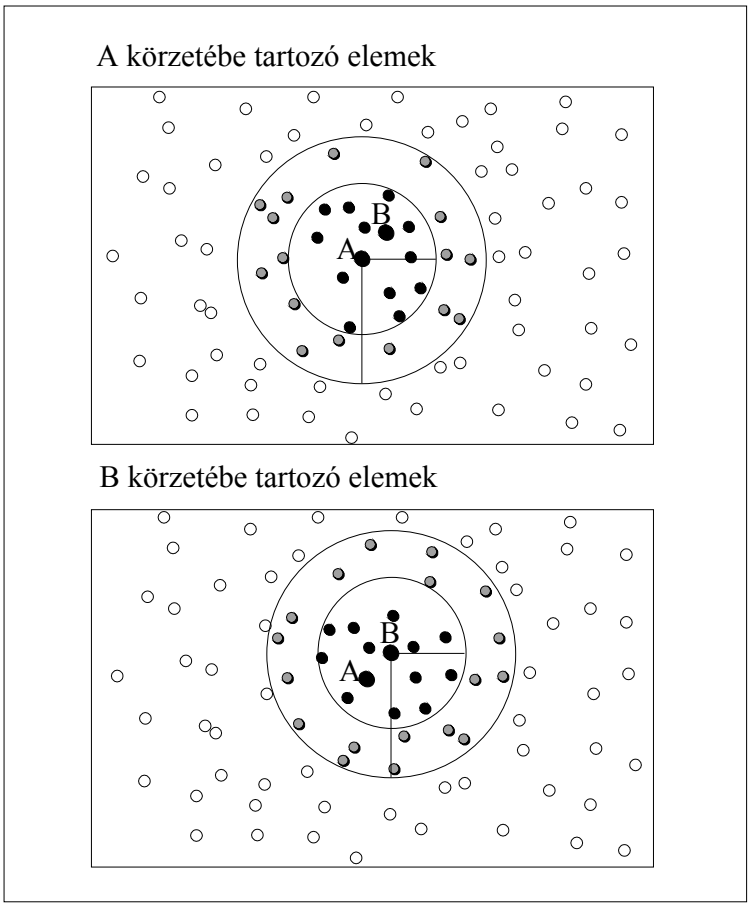
Egy adott elemi egységnél (leggyakrabban településnél) a vizsgált jellemző területi mozgóátlagát úgy számítjuk ki, hogy a jellemzőnek a településre és (alapesetben) attól *adott távolságon belüli* településekre rendelkezésre álló értékeit átlagoljuk (fajlagos adatok esetében *súlyozott* átlaggal számolva):

$$M(x_i)= \Sigma f_j*x_j/\Sigma f_j, \text{ azon elemekre, ahol } d_{ij}<=m$$

Az összefüggésben $M(x_i)$ x_i pont mozgóátlaga
d_{ij} i és j pont távolsága
m a mozgóátlag kiterjedése (leggyakrabban távolsági limit).

Az *2.6. ábrán* a legegyszerűbb esetet, vagyis pontalakzatot és légvonalbeli távolságot feltételezve látható a mozgóátlag kiszámításának logikája. Mindez kiterjeszthető térségi adatokra és bármilyen távolságfogalomra.

A módszer az összes elemi egységhez egy nagyobb térségre jellemző értéket rendel hozzá, végig haladva az összes elemi egységen. A számítás nem kötődik semmilyen már létező területi felosztáshoz (az elemi egységek lehatárolásától természetesen eltekintve). A módszer matematikai értelemben bármilyen területi bontásban elérhető mennyiségi ismérvre kiszámítható, azonban csak megfelelő területi részletezettségű adatok rendelkezésre állásakor érdemes alkalmazni. Magyarországon ez legalább kistérségi szintet jelent.



2.6. ábra *Az állandó sugarú mozgóátlag számítása kétfajta sugárral*

A területi mozgóátlag alaptípusai

- változatlan területnagyságú (a számítás egyszerű módja alapján állandó sugarú körűnek is nevezhető)
- változatlan súlyú (változó sugarú körrel)

Első esetben *azonos nagyságú területek* átlagát számoljuk, második esetben az átlagolandó terület nagyságát úgy határozzuk meg, hogy a súlyozási szempont szerint egyenlő súlyú körök jöjjenek létre. A súlyozási szempont leggyakrabban a népességszám, de az elemzés céljai szerint más is lehet. A változatlan területnagyságú mozgóátlagot számíthatjuk súlyozott és súlyozatlan formában is, a súlyozott forma az alapeset. A változó sugarú körök a vizsgálat fő szempontja szerint *azonos súlyú egységeket* hoznak létre (például azonos népességszámú, azonos nagyságú személyi jövedelem tömeggel vagy személygépkocsi állománnyal rendelkeznek a települések környezetébe eső települések).

2.7.2 A mozgóátlag kiterjedésének meghatározása

A települések egymástól való távolságának megállapításánál általános esetben a településközpontok *légvonalbeli távolságát* mérjük meg. A sugarak nagyságát az értelmes alsó és felső határ között rugalmasan határozhatjuk meg, annak megfelelően, hogy milyen területi szintű különbségeket szeretnénk vizsgálni. A számunkra érdekes területi szint adódhat valamilyen létező területi felosztás egységeinek átlagos nagyságából vagy egy intézmény, szervezet vagy gazdasági egység gyakorlatban megfigyelt átlagos vonzáskörzete alapján. Egy már létező területi felosztás (pl. a megyék) átlagos nagyságához igazodó igazodó sugár választása azért is előnyös, mert amennyiben az a településhálózat

bizonyos mértékű szabályosságaira épül, akkor az átlagolás a településhierarchia mentén meglévő ingadozásokat egyenlíti ki legnagyobb mértékben, és a térségi különbségeket emeli ki.

Az *azonos súlyú területek* súlymegállapítási elveiről hasonlóak mondhatóak el, mint az azonos sugarú körök sugarának kiválasztásáról. Fontos különbség adódik azonban abból, hogy az elemzési alapegységek, a területegységek már önmagukban is csoportosító ismérvekként funkcionálnak. Így a kör nagyságát fokozatosan növelve a változó nagysága nem fokozatosan növekszik, hanem egy-egy újabb területegység körön belülre kerülésével ugrásszerűen. A megállapított súlyt így szinte minden esetben a körök súlya vagy kicsivel alulmúlja, vagy meghaladja. A két érték közül azzal számolunk, amelyik közelebb esik a kívánt súlyhoz. Ez a probléma csak akkor vetődik fel élesen, hogyha számos olyan település létezik, amelyik önmagában is meghaladja a megállapított súlyt, más szóval ha túl kicsi a megállapított súlyunk a települések átlagos méretéhez képest.

Objektív kritériumunk és egyben megkötöttségünk tehát nincsen a mozgóátlag kiterjedésének (a területegység nagyságának vagy súlyának) megállapításánál, rossz kiterjedést nem választhatunk, legfeljebb jobban és kevésbé jobban indokolhatót.

2.8 Számítástechnikai alapok (Jakobi Ákos)

2.8.1 A területi elemzésben leggyakrabban használt szoftverek

A területi elemzés folyamatában rendszerint több szoftverrel is találkozik a felhasználó. Ha a legelső billentyűérintéstől a legutolsó lépésig értelmezzük az elemzés folyamatát, akkor igen sokirányú kapcsolatba kerülhetünk a számítógéppel. Ez egyfajta alapvető jártasságot, biztonságot igényel a komputer használatában, másrészt speciális tudást is a programok helyes futtatásában. A számítógéppel segített kutatás igazán jól tehát akkor fog működni, ha elegendő gyakorlattal rendelkezünk. Kezdő elemzők legnagyobb gondja az, hogy a szoftverhasználat során, tartva a helyrehozhatatlan következményektől, munkájukban már egy-egy kisebb problémánál is elakadnak. A tanulók gyakorlattal, segédanyagok felhasználásával e töréspontokon továbblendíthetők, ugyanakkor igen nagy szerepe van az *önálló problémamegoldásnak* is. Célszerű egy bizonyos jártassági szint után a részben önálló, majd teljesen önálló feladatmegoldásig eljutni, amely végső soron ahhoz vezet, hogy az elemzési problémák számítógépes megoldása már nem csak egyfajta feladatsor végrehajtása, hanem logikus gondolatok levezetése. Hangsúlyozzuk, hogy a számítási módszereket szinte mindenki – receptszerűen – végre tudja hajtani, de ez nem elég, *a végrehajtott vizsgálat elméleti alapjait is ismerni kell* a felmerülő problémák kezelése és a helyes értelmezés érdekében[8].

2.8.2 Szövegszerkesztés

A modern kor már mindennapos szinten is megköveteli, hogy írásainkat számítógéppel is szerkeszthető formában készítsük el. A külalaktól kezdve a szerkeszthetőségen át az elektronikus kommunikációban való továbbítási lehetőségekig számos előnyét tapasztalhatjuk a számítógéppel készített dokumentumainknak. Legtöbb esetben már a területi elemzéseket megrendelők is digitális formában kérik a munka végeredményeit, vagy legalábbis ebben a formában is a nyomtatott verzió mellett. Személyes versenyképességünk szempontjából ezért fontos lehet a szövegszerkesztő programok ismerete.

A szövegszerkesztő programok a területi kutatásokban az eredménymegjelenítésben és közzétételben játszanak fontos szerepet. Az effajta programok közül legismertebb *Microsoft Word* (MS Word) program, de egyéb szövegszerkesztő programok is léteznek (pl. *EasyWord*, *AbiWord*, *TextShield*). Nem szabad összekevernünk a szövegszerkesztést a kiadványszerkesztéssel, amely a már kész

[8] A területi elemzés számítástechnikai alapjainak jó összefoglalását tartalmazza *Herendi et al. 1996, 1-8. fejezet.*

szöveges és egyéb dokumentumok nyomdai szintű megszerkesztését jelenti. Ez utóbbi esetben már olyan feladatok is felmerülnek, melyek nyomdai alapismereteket igényelnek.

A Word program a területi elemzők számára leginkább az írógépet helyettesíti. Könnyű kezelhetősége lehetővé teszi, hogy a szöveges információkat gyorsan és tetszőleges formában meg tudjuk jeleníteni, továbbá hogy újrakezdés nélkül alapvető változtatásokat tudjunk elvégezni a szövegben. Gyakorlottabb felhasználók nem csak a szöveges információk beírására és rendezésére használják a programot, hanem komplett dokumentumok elkészítésére is. A program ugyanis lehetővé teszi táblázatok, ábrák, térképek, képek vagy speciális karakterek beszúrását is, ami a produktumok fogyaszthatósága szempontjából nem utolsó előny. A területi elemzők számára különösképpen fontos lehet az egyéb programokban kiszámított vagy létrehozott részeredmények beillesztése a dokumentumba, mondjuk a bekezdések közé szúrva, ami viszont nagyfokú jártasságot igényel a *programok közötti átjárhatóság* szempontjából.

A Word program alapvető fájlformátuma a Word dokumentum (.doc), ugyanakkor számos klasszikus formátumú fájl is megnyitható a programban (például .txt). Abban az esetben, ha a kibocsátó és a cél-számítógép eltérő szövegszerkesztő programmal rendelkezik, vagy ha ismeretlen szoftverrel rendelkező gépre szeretnénk átvinni digitális szöveges dokumentumunkat, érdemes az ún. Rich Text Format (.rtf) formátumú verzióban elmentenünk anyagunkat, amely – a bonyolultabb betéteket leszámítva – nagy eséllyel ugyanazt a képet fogja nyújtani a befogadó gépen, mint a sajátunkon.

Az időről időre piacra kerülő új szoftver verziók kompatibilitási gondokat okozhatnak a korábbi verziók működésében. A modernebb programokkal készített fájlok esetenként olyan újításokat is tartalmaznak, amelyek egyszerű technikai okoknál fogva nem értelmezhetők a régebbi verziókban. Ügyelnünk kell tehát arra, hogy egy-egy program melyik verziójával dolgozunk, és mely verziójában szándékozunk elmenteni végeredményeinket. A modernebb programok általában ismerik a korábbi verziók formátumait, visszafelé ez azonban nem oly gyakran igaz: ez a *lefelé kompatibilitás szabálya*.

Mivel a szövegszerkesztők általában az eredmények írásos megjelenítésére használatosak, általánosságban jellemző az egyéb programokban elkészített részeredményeknek e programokba való átvitele, s csak kisebb arányban a szövegszerkesztőkben létrehozott eredmények átvitele más szoftverekbe. Az eredmények exportálásának és importálásának legfontosabb módjait a 2.*4. táblázatban* közöljük (MS Word 2000-re vonatkozóan).

Importálás: Korábban már létező anyagok megnyitása (Fájl / Megnyitás)	**Beszúrás:** Részeredmények beillesztése a szövegbe (Beszúrás...)	**Exportálás:** Megszerkesztett anyagok kivitele a programból (Fájl / Mentés másként)
Word dokumentum (.doc) Rich Text formátum (.rtf) Szövegfájl (.txt) Word Perfect (.doc) Word 6.0/95 (.doc) Word 2.x (.doc) Works 4.0 (.wps)	raszteres képfájlok (.gif, .jpeg, .jpg, .tif, .bmp) vektoros képfájlok (.wmf, .cdr) MS Excel munkalap (.xls)	Word dokumentum (.doc) Rich Text formátum (.rtf) Szövegfájl (.txt) Word Perfect (.doc) Word 6.0/95 (.doc) Word 2.x (.doc) Works 4.0 (.wps) Weblap (.htm)

2.4. táblázat *Segédlet a programok közötti átjáráshoz (MS Word 2000)*

2.8.3 Táblázatkezelés

A területi vizsgálatok leggyakrabban alkalmazott beviteli formáit a területi adatmátrixokban összefoglalt adathalmazok képezik. A területi kutatóknak azonban nem csak az adatbevitel kapcsán lehet szüksége táblázatkezelő programokra. A legtöbb ilyen program tulajdonképpen statisztikai elemző programként is funkcionál, bár az adatelemzésre kimondottan „szakosított" programok is léteznek.

Bár egyszerű számításokat a WORD szövegszerkesztőbe illesztett táblázatok adataival is végezhetők (*Táblázat/Képlet* menü pontok), a táblázatkezelő programok közül legelterjedtebb és legismertebb a *Microsoft Excel (MS Excel)* program, amely mind a számításokat, mind a táblázatformázás műveleteit lehetővé teszi. A program népszerűsége egyrészt logikus kezelhetőségének, másrészt széleskörű felhasználhatóságának is köszönhető. Használják például nyilvántartási rendszerekben, adatbázisok készítésekor, egyszerűbb és bonyolultabb számítási műveletekben stb. A program kimeneti „termékei" lehetnek például maguk az összeszerkesztett táblázatok, adathalmazok, továbbá a számítási eredmények és nem utolsó sorban grafikus produktumok is (grafikonok, diagramok).

Importálás: Korábban már létező anyagok megnyitása (Fájl / Megnyitás)	**Exportálás:** Megszerkesztett anyagok kivitele a programból (Fájl / Mentés másként)
Excel Munkafüzet (.xls) Excel Munkalap (.xls) Excel Munkaterület (.wks) dBASE (.dbf) Quattro Pro (.wq1) Excel 5.0/95 (.xls)	Excel Munkafüzet (.xls) Excel Munkalap (.xls) Excel Munkaterület (.wks) dBASE (.dbf) Quattro Pro (.wq1) Excel 5.0/95 (.xls) Excel 4.0 munkalap (.xls) Data Interchange (.dif)

2.5. táblázat *Segédlet a programok közötti átjáráshoz (MS Excel 2000)*

Az Excel program saját fájltípusai az Excel Munkafüzetet vagy Munkalapot tartalmazó (.xls) fájlok illetve az ún. egész Munkaterületet tartalmazó (.wks) fájlok. A program általában képes más táblázatkezelő programok által készített fájlokkal is dolgozni, melyek közül legismertebbek a dBASE-fájlok (.dbf), illetve a Quattro Pro (.wq1) fájlok, másrészt az itt elkészített adatfájlok alapjául szolgálhatnak más, magasabb szintű műveleteket is elvégezni képes programoknak.

2.8.4 Adatelemzés

Maradjunk még kicsit az *Excel* programnál és a területi adatmátrixoknál. E programnak a területi elemzők számára egy igen közkedvelt funkciója a számítások, egyenletek, függvényműveletek stb. egyszerű és gyors végrehajtása a megfelelően összeállított adatbázisokon. Helyesebben fogalmazva az egyszerűtől a bonyolultig sok adatelemző, adattranszformáló vagy statisztikai funkció érhető el, amelyek a regionális elemzési módszerek számítógépes kivitelezéséhez igazán nagy segítséget nyújtanak. Az Excel tekinthető leginkább annak a szoftvernek, amellyel a területi elemzéseink legnagyobb része számítógépen elvégezhető, mindazonáltal ez a program sem tud mindent.

A program hátránya – mások szerint előnye – hogy többnyire nem kész modulokkal rendelkezik a különféle számítások elvégzéséhez, hanem nekünk, elemzőknek kell a számítás lépéseit, egyenleteit megadnunk. Nincs tehát szó arról, hogy a begépelt vagy importált adathalmazunkból egy kattintással kiszámolja a program a kívánt eredményeket. A legegyszerűbb vizsgálati módszerektől eltekintve logikus lépések sorából felépülő számítássorozatot kell általában elvégeztetnünk a programmal. Mindez tulajdonképpen abból fakad, hogy ez a program egyszerre csak egy számítási utasítást tud

végrehajtani, aminek viszont megvan az az előnye, hogy a számításnak minden lépését koordinálhatjuk, felügyelhetjük vagy akár módosíthatjuk.

A társadalomkutatók és területi elemzők számára egyik legfontosabb program az adatelemző szoftverek csoportján belül az *SPSS programcsomag (Statistical Package for Social Sciences = Társadalomtudományi Statisztikai Programcsomag*). A szociológiában, a közgazdaságtanban, a statisztikában, a pszichológiában, a biometriában illetve egyre inkább a területi elemzésekben is alkalmazott programcsomag (verziótól függően) ablakos-menüs rendszerben könnyű kezelhetőség mellett teszi lehetővé a legbonyolultabb és legmagasabb szintű elemzési módszerek végrehajtását. Helyes alkalmazásához a technikai szempontú gyakorlat mellett az elméleti-statisztikai háttér magas fokú ismerete is szükséges.

Adatbevitel: Korábban már létező anyagok megnyitása (File / Open)	**Eredménykivitel:** Megszerkesztett anyagok kivitele a programból (File / Save as)
SPSS tábla (.sav) SPSS/PC régi verz. (.sys) Excel 4.0 munkalap (.xls) dBASE (.dbf) Tab-delimited (.dat, .txt) SPSS Syntax (.sps) SPSS Viewer (.spo)	SPSS tábla (.sav) SPSS/PC régi verz. (.sys) Excel 4.0 munkalap (.xls) dBASE (.dbf) SPSS Syntax (.sps) SPSS Viewer (.spo) HTML file (.htm) Text file (.txt)

2.6. táblázat *Segédlet a programok közötti átjáráshoz (SPSS 8.0)*

Az SPSS programcsomag leginkább az összetett és többváltozós vizsgálati módszerek végrehajtásában nyújthat segítséget. Számos olyan funkciója létezik, amely más programokkal nem kivitelezhető, illetőleg rendkívül hosszadalmas számításokkal helyettesíthető. Célszerű tehát akkor nyúlni ehhez a programhoz, amikor már tisztában vagyunk a vizsgálati módszerek egyszerűbb változataival.

Az SPSS-sel, illetve a hasonló programokkal elvégzett elemzések általános menete a következő lépésekben fogalmazható meg: 1. Az adatok bevitele, importálása vagy begépelése az adatszerkesztő felületre, 2. A vizsgálati módszer kiválasztása a menürendszerből, 3. A változók kiválasztása az analízishez, 4. Az eredmények értékelése. Bár a program adatelemző szoftverként ismert, a számítások során kiszámított és megjelenített eredményeket mégsem elemzi le „teljes mértékben". A végeredményként kapott újabb táblákat és számokat – értelemszerűen – már nekünk kell értékelnünk. A gép csak az eljárás végrehajtását könnyíti meg, az elemzést az ember végzi!

Az SPSS program mellett más adatelemző szoftverek is megemlíthetők, melyek főként a statisztikai elemző módszereket támogatják (pl. Statistica, Stata). E programok bizonyos elemeikben eltérő eljárásokat is végre tudnak hajtani (pl. ún. panel-számításokat vagy trendfelület-vizsgálatokat).

2.8.5 Prezentációs eszközök

A prezentáció, azaz az eredmények bemutatása a területi elemzéseknek nem, az elkészített anyagok közzétételének azonban lényeges eleme.

A kész produktumok és kutatási eredmények nem írásos formájú bemutatása, közzététele sajátos eszközöket igényel. Az eredmények előadás formájában történő bemutatásához nagy segítséget nyújtanak az olyan speciális előadás-tervező szoftverek, mint például a *Microsoft PowerPoint* programja. Ez az egyszerűen kezelhető, rövididejű gyakorlást igénylő szoftver változatos és látványos formában képes bemutatni eredményeinket. E program népszerűsége részben könnyű kezelhetőségéből ered, másrészt egyre jelentősebb elterjedtségéből a piaci szférától az oktatási és tudományos körökig.

E program segítségével az eredményeink szóbeli ismertetésével párhuzamosan írásbeli és képi összetevőket is megjeleníthetünk. A jó PowerPoint bemutató vázlatpontokba sűrítve jeleníti meg mondanivalónkat, amelyet az előadásunkban fejtünk ki bővebben és nem a bemutatásra kerülő anyagban. A bemutató tulajdonképpen általunk megszerkesztett diakockák kivetítéséből áll, amely felfogható egymás után következő írásvetítő-fóliáknak is. A megjelenítéshez is egy írásvetítőhöz hasonló eszközre, ún. projektorra van szükség.

Bevitel: Korábban már létező anyagok megnyitása (Fájl / Megnyitás)	**Beszúrás:** Részeredmények beillesztése a szövegbe (Beszúrás...)	**Kivitel:** Megszerkesztett anyagok kivitele a programból (Fájl / Mentés másként)
PowerPoint bemutató (.ppt) PowerPoint vetítés (.pps)	raszteres képfájlok (.gif, .jpeg, .jpg, .tif, .bmp) vektoros képfájlok (.wmf, .cdr) MS Excel munkalap (.xls)	PowerPoint bemutató (.ppt) PowerPoint vetítés (.pps) raszteres kép (.gif, .jpeg, .tif) vektoros kép (.wmf) szöveges vázlat (.rtf)

2.7. táblázat *Segédlet a programok közötti átjáráshoz (MS PowerPoint 2000)*

2.8.6 Speciális szoftverek

Léteznek a területi elemzők által használt szoftverek között olyanok is, amelyek egy-egy speciális elemzési módszerhez nyújtanak segítséget, de a területi elemzésekben általános szerephez nem jutnak. Ezek a szoftverek rendszerint más, nem területi elemzési céllal készültek, de bizonyos esetekben kifejezetten alkalmasak egy-egy módszer számítógépes végrehajtására.

Elsőként említsük meg a térinformatikai, digitális képfeldolgozási és távérzékelési elemző eljárásokhoz alkalmazott *Idrisi* programot, amely jelentős segítséget nyújthat összetett légi- és űrfelvételek szakszerű, objektív és gyors elemzéséhez. A program viszonylag elterjedtnek számít a térképész, földmérő stb. szakmákban, de a területi elemzők is hasznát vehetik például a tájértékelés vagy az alakzatvizsgálatok esetében. A program az adatelemzés mellett a térképi adatok megjelenítését is támogatja.

Ugyancsak többségében térinformatikusok használják, de a területi elemzésekben is szerephez juthat a *Surfer* program, amely két- és háromdimenziós felületek készítésére alkalmas. Kedvelt alkalmazási területe a digitális terepmodellek készítése és térbeli megjelenítése. A területi elemzők számára a trendfelület-elemzésben és a társadalomföldrajzi domborfelületek készítésében lehet jelentősebb szerepe.

A *VisualRoute* nevű program alapvetően informatikai és hálózatelemzési céllal készült, de a leleményes területi kutatók e programban is felfedezték a számukra használható elemeket. A program segítségével a *kiberteret,* azaz a világháló belső térbeli kapcsolatait, hálózati összeköttetéseit tudják vizsgálni és térképezni.

Általánosan elterjedtek azok a szoftverek, amelyek a területi elemzések mellékleteiként felhasználásra kerülő ábrákat képesek elkészíteni. Ábra lehet természetesen bármely, az előbbiekben bemutatott szoftver végterméke is, ugyanakkor elképzelhető, hogy az elkészíteni kívánt ábránk nem illeszthető egyetlen programhoz sem. A speciális tematikus térképektől kezdve, a vázlatokon át a magyarázó tartalmú ábrákig sok esetben kerülhetünk olyan helyzetbe, amikor a rajzoló vagy illusztráló programok nyújtják az egyetlen kiutat. Az egyik legnépszerűbb ilyen program a már említett *Adobe Photoshop* program „párja”, az *Adobe Illustrator,* amely egyszerre képes raszteres és vektoros alapon dolgozni. Hasonlóan igaz ez az állítás a könnyen kezelhető *Freehand* programra, illetve a professzionális kivitelezést is lehetővé tevő és igen elterjedt *Corel Draw* programra (és kiegészítőire).

2.9 Grafikonok (Szabó Pál)

A *vizualizáció* generális elemzési és prezentációs eszköz. Sokak számára a különböző ábrák jóval érthetőbbek, mint az ugyanazon jelentést hordozó száraz számsorok. A területi kutatásokban egyrészt *a statisztika általános grafikus módszerei*, másrészt – a tudományterület talán legsajátabb eszközeként – a különböző *térképezési eljárások* kiemelt fontosságúak.

A *grafikus megjelenítés azonban soha nem helyettesíti az elemzést,*. az adatok értékelését és megfontolandó az is, hogy mennyiben élünk az ábrakészítés lehetőségével, mivel ha túltengenek a diagramok, grafikonok egy munkában, könnyen elvonhatja a figyelmet az érdemi információkról. Ez utóbbi hibát tetézi, ha a vizuális élményadást azzal fokozzuk, hogy a legkülönbözőbb ábratípusokat tömegesen variáljuk. Ez az út nem a kutatási célt szolgálja. Összességében soha nem szabad arról megfeledkeznünk, hogy *a grafikus ábrázolás a kutatási munkában eszköz, és nem cél*.

2.9.1 Általános szempontok

A következőkben összefoglaljuk a területi kutatásokban használatos és használható diagramok és grafikonok készítésének általunk legfontosabbnak ítélt általános szempontjait (részben *Hunyadi L. 2002* alapján), a részletekre a konkrét eseteknél, a diagram- és grafikontípusok bemutatásánál térünk ki.

- Az ábra legyen *célorientált*, azaz a bemutatásra szánt információkat tartalmazza, s ne zsúfoljunk tele egyéb információkkal. Kerüljük a bonyolult, összetett, többféle célt egyszerre szolgálni akaró ábrázolást.

- Az ábra legyen *egyszerű és áttekinthető*, azaz a befogadó rövid szemlélés alapján is fel tudja dolgozni a grafikus formában nyújtott információkat. Kerülni kell például a felesleges formai (pl. háromdimenziós kúpok), színbeli (pl. halacskás tapéta a háttérben) extrákat, illetve az aprólékos információközlést (pl. kategóriatengelyen rendkívül sűrű számskála). A külcsín, a sok – gyakorta felesleges – felirat stb. elvonhatják a figyelmet a lényegi közlésről.

- Az ábrán *minden információ rajta legyen*, de se több (ne legyen ismétlődések), se kevesebb (egyetlen lényeges információ se maradjon le). A legszükségesebbek az ún. alapinformációk (az ábratípustól függően a következők valamilyen kombinációja): terület – vizsgált terület, területi keret, területi egységek; jelenség – a vizsgált jelenség, mutató(k), mértékegység(ek); idő – időpont(ok), időintervallum.

- Az ábra *optikailag semleges legyen*, ne manipuláljuk vele a befogadót. Ez vonatkozhat például a skálázásra: kerülni kell, illetve óvatosan kell bánni a koordinátarendszer tengelyei metszéspontjainak meghatározásával, illetve gondosan kell megválasztani, a céltól függően, a skálák minimumát, maximumát. (Például különböző maximum- ill. minimumértékek ugyanazon adatsor esetén csökkenthetik vagy növelhetik az adatok közötti különbségek látszatát.) Kerülendők a háromdimenziós diagramok, mivel már a síkbeli formáció is torzíthatja a számértéket: nem mindegy, hogy az alakzat térfogata, területe vagy az idom magassága egyenlő az ábrázolt adattal.

- *Az ábrának legyen címe*. A cím lehetőleg ne legyen hosszú. A címben az alapinformációk jelennek meg: általában szerepel a vizsgált területi keret (esetleg a területi bontás), a jelenség (vagy annak konkrétan a mutatója) és a vizsgált időszak, időintervallum (de például az ábrázolt mutató mértékegységét nem írjuk bele a címbe). Nem igazán jó cím „Egy főre jutó GDP, 2000”, szemben „A fejlettség regionális egyenlőtlensége Belgiumban, 2000” címmel, illetve „Népesség

(1950-2000)", szemben „A népességszám változása Magyarországon a XX. sz. második felében" címmel.

- Általában szükséges *jelmagyarázat* az ábrához. Ez világos, egyértelmű legyen, és ne olyan információkat adjon, melyek már szerepelnek más formában az ábrán.

- Az ábra *színezésénél óvatosan járjunk el*. Egyrészt célszerű a fekete-fehér ábrák használata (esetleges nagyszámú nyomtatás, fénymásolás stb. miatt), de ha mindenképpen színes ábrát készítünk, akkor is figyelni kell a különböző színek összeválogatására. A diagramoknál az alakzatok (oszlopok, körcikkek stb.), a grafikonoknál a pontok színessé tételénél kerüljük a „nem természetes" (neonzöld stb.) és a rosszul látható (pl. halvány) színeket, másrészt az egyes színek, színárnyalatok szerinti megkülönböztetésnél pedig a különbségtételnek határozottnak kell lennie (mind az ábrán, mind a jelmagyarázatban szereplő színek alapján egyértelműen kell tudni azonosítani az egyes egységeket). A színválasztásnál érdemes elgondolkodni azon, hogy az adott jelenséghez nincs e valamilyen szín konvencionálisan társítva (pl. mezőgazdaságnál a zöld), mert az ettől elütő színválasztás nem mindig szerencsés. Szintén érdemes odafigyelni, hogy a színskála (pl. kördiagramnál) ne össze-vissza legyen, hanem a tematikus térképezés elvét követve (pl. hőskála) sorakozzanak egymásután a színek. Magának az ábrának a hátterét általában fehérre érdemes „színezni", illetve olyanra, mely a legélesebb kontrasztot adja az ábra többi részével. Nem érdemes felesleges extrákkal fűszerezni a hátteret, legfeljebb az ábrázolt jelenséget szimbolizáló, egyszerű és halványan megjelenő képpel.

- A színezés mellett a *sraffozásos* ábra (amikor pontok, vonalak, rácsok stb. különítik el az egyes egységeket) is elterjedt, azonban ennél több megszívlelendő tanács is van. Egyrészt a jelmagyarázatban, szerény méretű szimbólumnégyzet esetén, nem mindig kivehető a sraffozás típusa (pl. ritka vonalak, ritka pontok), s így a beazonosítás gondot okoz. Másrészt célszerű követni a tematikus térképezés szabályait, konkrétan a „sűrűsödés" elvét, azaz az egymás után következő (pl. körcikkek) egyre sűrűbben legyenek kitöltve (pl. fehér szín, pontok, sűrű pontok, vonalak, sűrű vonalak, rácsozat, fekete).

- Jó, hogy ha az ábra *önállóan (a szövegtől függetlenül) is megállja a helyét*. Gyakorta előfordul, hogy az ábrát más szerzők is felhasználják munkáikban, s ilyenkor értelmét vesztheti az ábra tartalma, ha hosszadalmas leírás párosult hozzá az eredeti szövegben.

- Általában szükséges a *feldolgozott adatok forrásának feltüntetése*, a *„forrás: ..."* vagy a *„saját szerkesztés ... forrás adatainak felhasználásával"* információ felvitele az ábrára, átvett információk híján, a teljesen önállóan készült ábráknál természetes a *„saját szerkesztés"* feltüntetése felesleges! Ha több ábra esetében azonos a forrásháttér, elég a legelső esetben vagy a szövegben jelezni.

- A végére hagytuk az egyik leglényegesebb elvet: mindig *a jelenséghez rendelhető ábratípust válasszunk*. Az ábrákat nagyobbrészt területi (országok, régiók, települések stb.) adatokra készítjük, ami azonban korlátokat is jelenthet. Egyes esetekben többféle típus is használható, de vannak olyan statisztikai adatsorok, melyeknél nem. Például egyes fajlagos adatokat (népsűrűség, egy főre jutó GDP stb.), mivel nincs értelme a (súlyozatlan) összegüknek, nem ábrázolunk olyan diagramon, ahol az értékek összegeződnek (pl. kördiagram, halmozott oszlop).[9]

A következőkben a *legegyszerűbb grafikus ábratípusokat* vesszük sorra. Minden esetben az általánosan elterjedt, magyar nyelvű EXCEL programmal készült példákat mutatunk be, kiemelve a területi kutatások keretén belül készíthető ábrákkal kapcsolatos jellemzőket, problémákat.

---

[9] A szerzők naív óhaja a kézirat leadásakor: talán kötetünk ábrái e szigorú kritériumok többségének megfelelnek majd!

A különböző diagramok, és grafikonok közül EXCEL-ben a Beszúrás/Diagram menüpontból illetve a DiagramVarázsló ikon aktiválása után tudunk választani.

Az EXCEL-ben készített ábrák mindig „igazodnak” az eredeti adattáblához, így az azon történő bármilyen változtatást automatikusan követi az ábra. Ennek előnye, hogy a tévesen bekerült adat(ok) korrigálása után nem kell újból elkészíteni a diagramot, grafikont, hátránya viszont, hogy az adattáblán végzett műveletek megváltoztatják az ábrát. Emiatt célszerű mindig külön táblát készíteni az adatokból az ábra „alapanyagául”, és ezt a továbbiakban, ha csak nem szükséges, akkor nem bolygatni. Szintén az EXCEL program jellemzője, hogy a többlépcsős ábrakészítés bármelyik lépcsőfoka újra behívható és módosítható, azaz egy kész ábrán is minden újból át- illetve beállítható.

2.9.2 Oszlopdiagram

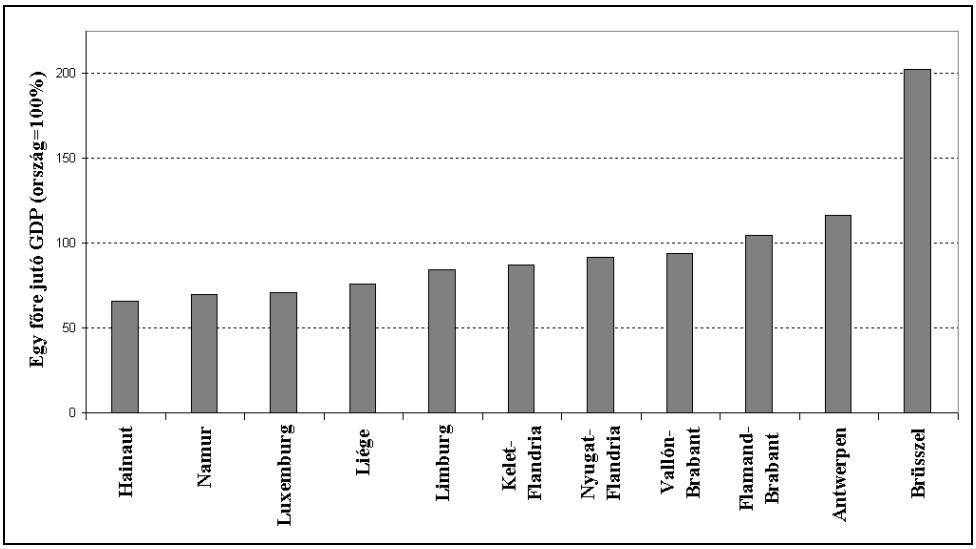
A diagramot leginkább arra használjuk, hogy egy vagy több területi adatsort ábrázoljunk, oly módon, hogy *egy területegység értékét a hozzá rendelt oszlop Y tengelyen leolvasható magassága adja meg*. A diagramot mind abszolút, mind relatív mutatókra elkészíthetjük.

Az oszlopdiagramoknak többféle altípusa van:

- A *csoportosított oszlop* típust alkalmazhatjuk mind abszolút, mind relatív mutatóknál például a különböző térségek azonos mutatójú adatainak összehasonlítására (*2.7. ábra*).

Néhány további (nagyon) gyakorlatias szempont:

Túl sok térséget nem célszerű így ábrázolni, mivel nehezen kibogozhatóvá válik az ábra, túl kevésért pedig nem érdemes külön ábrát készíteni. A gyakorlat maximum kb. 20 egységet, minimum 4 egységet sugall. Lényeges kérdés ennél a típusnál az oszlopvastagság. Mivel az oszlop magassága (és nem területe) jelenti az adott értéket, emiatt nem érdemes túl széles oszlopvastagságot beállítani, mivel vizuálisan csökkenti a különbséget. Korrektek a szakaszok lennének, melyek hossza megegyezik az ábrázolt értékekkel, azonban ezek kevésbé kifejezőek a látvány szempontjából. Ide tartozik még az oszlopok összeérésének kérdése: összeérést csak akkor célszerű alkalmazni, ha van kapcsolat a szomszédos oszlopok között, például egy terület egy jelenségének időbeli változását mutatja az ábra. Figyeljünk arra, hogy minden oszlophoz tartozzon az X tengelyen azonosító, ne maradjon le például egyetlen régiónév se. Ezt hosszú nevek esetén úgy tudjuk elérni, hogy vagy kisebb betűre állítjuk át a szöveget, vagy elforgatjuk a neveket. Az Y-tengelyen szerepelnie kell az ábrázolt jelenség mutatójának, mértékegységgel együtt. Ez alól kivételt jelent, ha a mutató már magában a címben is előfordul, ekkor itt csak a mértékegységet kell feltüntetni. Az Y tengely maximumát (minimumát) úgy érdemes megválasztani, hogy a legnagyobb (legkisebb) érték is látszódjon. Ez alól kivételt jelent, ha egy messze kiugró értékkel rendelkező egységünk is van: ilyenkor „elvágjuk” az oszlopot (kisebb skálamaximum) és az értéket ráírjuk a „maradékoszlop” tetejére (vagy „megtörjük” a tengelyt). Értékeket általában nem szokás ráírni az oszlopokra, mivel a diagramnak nem az adatközlés a célja (arra az értékek szerint sorrendbe rendezett adattábla a célszerű), hanem a nagyságrendi megkülönböztetés. Lehetőség van vezetőrácsok beállítására. Ezeket akkor érdemes alkalmazni, ha túl sok egység áll közel azonos szinten, s így például a csoportképzésnél (magas, közepes, alacsony értékűek elválasztása) a konkrétan kijelölt értékhatárok szerint nehéz a besorolás. Ha egy vizsgált térségünk van, melyet egy tágabb keretben helyeztünk el, akkor a vizsgált egységet célszerű megkülönböztetni a többitől, például az oszlopot más színűre állítjuk át.

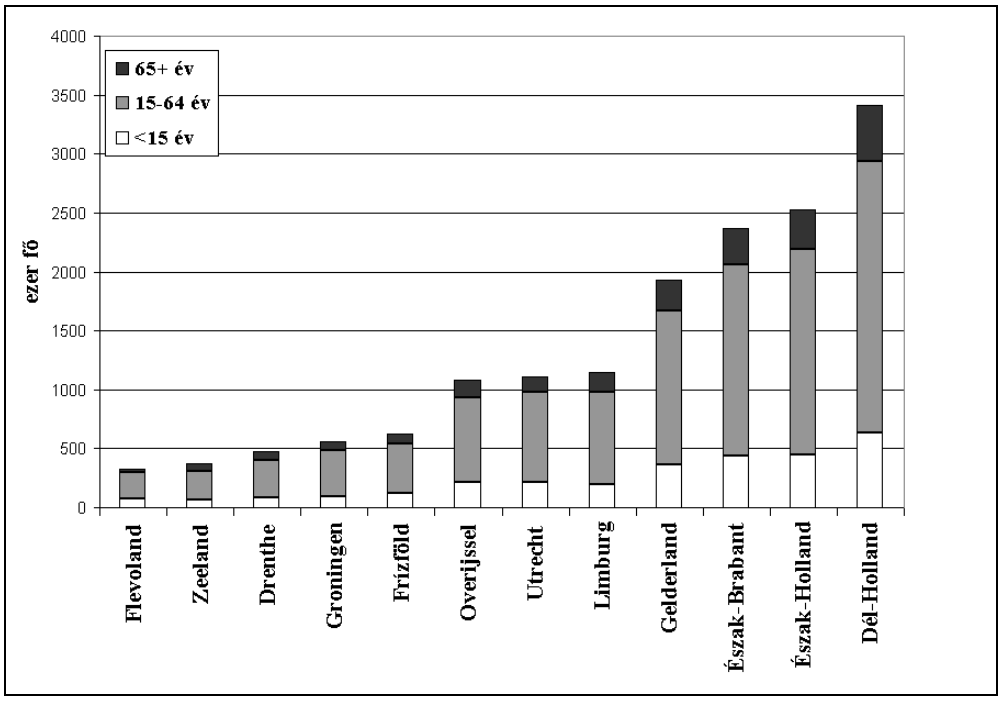


2.7. ábra *A gazdasági fejlettség regionális különbségei Belgiumban (2000)*

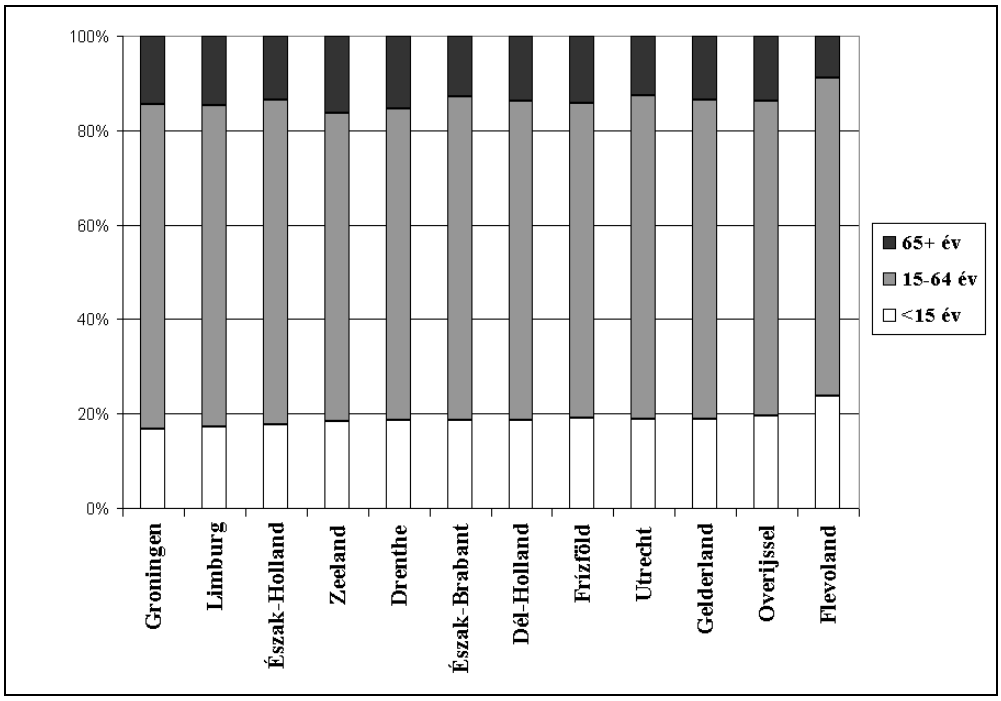
Ésszerű, hogy az X tengelyen sorakozó egységek valamilyen szempont szerint sorrendbe legyenek téve: ez leggyakrabban a mutató értékei szerinti (több mutatósor esetén az egyik, például a kiemelt mutató szerinti), de néha ettől eltérő szempontok is szóba jöhetnek (például földrajzi égtáj vagy népességnagyság stb.). *Mindenképpen legyen logika az X tengelyen elhelyezett egységek sorrendjében.* Leggyakoribb hiba a térségek – adattáblákban gyakorta használt, s így a diagramokon is megjelenő – ABC-sorrendben történő sorakoztatása. Ezt mindenképp kerüljük, emiatt azt ábra elkészítése előtt rendezzük sorrendbe a területegységeket, például az ábrázolandó mutató értéke szerint. Ebben az esetben gyakorlatilag mindegy, hogy csökkenő vagy növekvő sorrendbe következnek egymás után az egyes oszlopok.

Alkalmazható ez a diagramtípus egyszerre több mutató ábrázolására is, azonban ilyenkor problémát okoz a különböző skála (értéknagyságrend, mértékegység). Ebből kifolyólag általában csak az azonos karakterű mutatók együttes ábrázolása a célravezető, illetve a különböző mutatók azonos nevezőre hozása (például az országos átlag százalékában fejezzük ki az egyes értékeket).

- A *halmozott oszlop* diagramtípust több mutató egyszerre történő ábrázolásánál alkalmazhatjuk, azonban csak akkor, ha értelmezhető a különböző mutatók értékeinek összege. Az ábrán az egyes területegységekhez tartozó adatok összegződése történik meg (*2.8. ábra*).
- Ennek altípusa a *100%-ig halmozott oszlop*. Ilyenkor nemcsak összeadódnak az értékek, hanem a teljes százalékában fejezzük ki az egyes összetevőket (*2.9. ábra*). Ennél az ábratípusnál a fő gondot az jelenti, hogy több mutatót ábrázolunk egyszerre, s ilyenkor csak az egyik jelzőszám szerint tudjuk az egységeket sorba rendezni, s így csak ennek különbségei elemezhetők gyorsan és egyszerűen.



2.8. ábra A népesség korcsoport szerinti megoszlása Hollandia régióiban (2000)

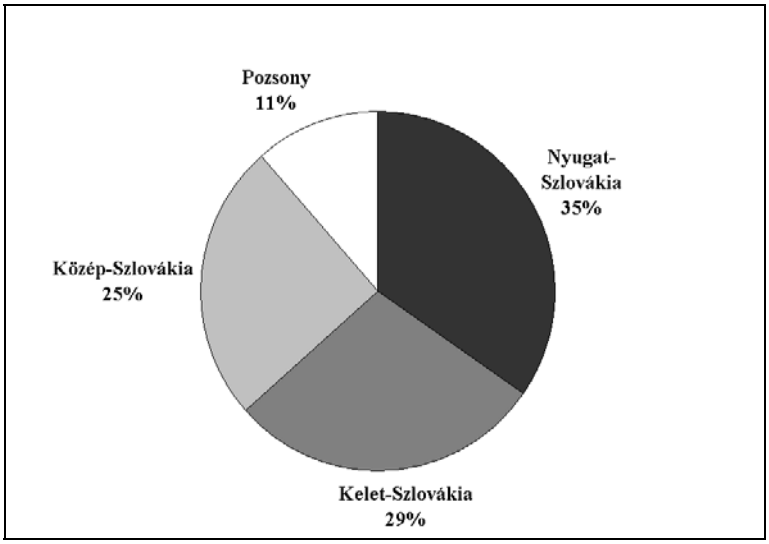


2.9. ábra A népesség korcsoport szerinti összetétele Hollandia régióiban (2000)

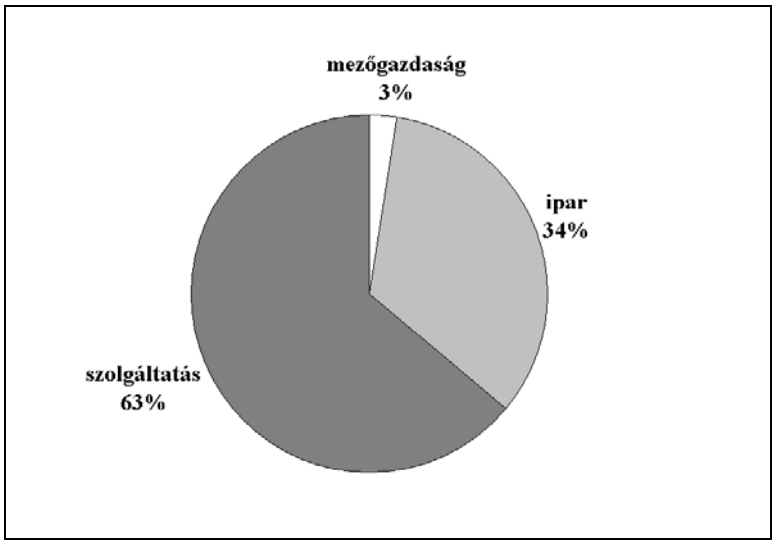
Az oszlopdiagramok párja a ritkábban használt *sávdiagram*, ahol vízszintesen helyezkednek el az egyes oszlopok. Jelentősebb különbség nincs a két típus között, s így gyakorlatilag a korábban felsorolt javaslatok az irányadóak. Külön csoportot alkotnak a téglalap helyett egyéb, általában háromszög, illetve a háromdimenziós típusoknál a henger, kúp, piramis alakzatok. Használatuk mellőzése javallt!

2.9.3 Kördiagram

Az oszlopdiagram mellett a másik legelterjedtebb diagramtípus. Lényege, hogy egy területi adatsor esetén az egyes területegységek részesedését adja meg a teljes területből, emiatt csak abszolút mutatókra illetve az ezekből számított százalékos értékekre készíthető (ez utóbbit az EXCEL program automatikusan megadja a diagram készítésénél). *Egy területegység értékét a hozzá rendelt körcikk területe (illetve ívhossza) adja meg* (*2.10. ábra*).



2.10. ábra *Szlovákia népességének regionális megoszlása (2000)*



2.11. ábra *Foglalkozási szerkezet Németországban (2000)*

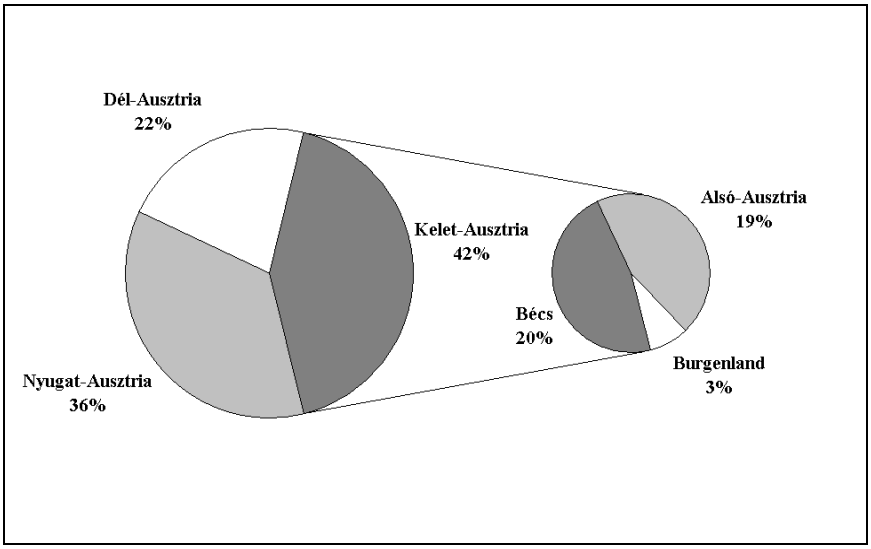
Gyakori használatával találkozhatunk még egy terület adatának (pl. foglalkozási szerkezet) felbontásakor (pl. primer, szekunder, tercier szektorban dolgozók aránya), ilyenkor az egyes körcikkek az adott jelenség részeinek értékét mutatják (*2.11. ábra*).

Az ábrázolt szektorok száma minimum 3 (két számra nem érdemes diagramot „pazarolni”), maximum 5-7 legyen. Több adat ábrázolása a kördiagrammon ugyan elvileg nem tiltott, ennél több egység esetén már gyakran nehezen értelmezhető az elaprózódó körcikkekre tagolt ábra, ilyenkor jobb az eredeti

adatokat táblázatban közölni, ha mindenképp szükségesnek tartjuk, hogy a kis értékű, súlyú területegységekről is pontos számszerű információt közöljünk.

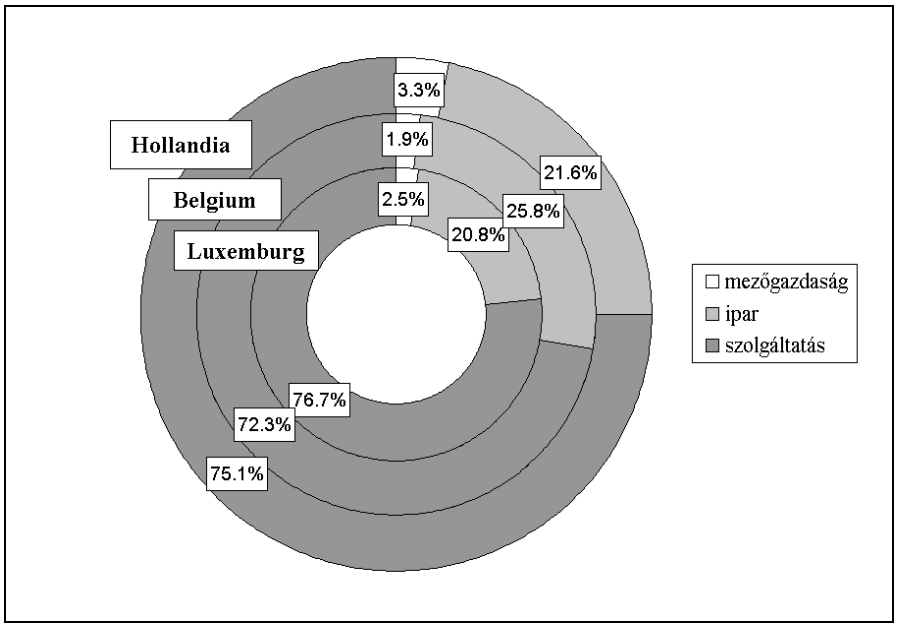
A kördiagramnál szintén célszerű a sorba rendezés, legjobb ha a cikkek csökkenő nagyságban követik egymást (lehet növekvő is a sorrend). Automatikusan 12 órától indul a kör (ezt, bár lehet, nem érdemes átállítani), és az óramutató járásával megegyező irányba követik egymást az egyes körcikkek. Ennél a típusnál, mivel a hangsúly a részesedésen van, jó ha feltüntetjük a százalékos (nem pedig az abszolút mutató szerinti) értéket. Emellett a körcikk „tulajdonosa" vagy a körcikk mellett szerepeljen (így az egyszerűbb az ábra értelmezése), vagy, ha túl kicsik egyes körcikkek, külön jelmagyarázatban (ez utóbbi esetben, habár nem mindig szerencsés, a jelmagyarázat kikerülhető, ha a cikk és tulajdonosa szakasszal van összekötve). A színskálára az előző alfejezetben leírtak a mérvadóak. Lehetőség van rá, hogy a körcikkeket széthúzzuk (robbantott kör), de ez csak zavarólag hat, mivel a körcikkek egymáshoz való viszonyítását optikailag megzavarja. De ha egy vizsgált térségünk van, melyet egy tágabb keretben helyeztünk el, akkor a vizsgált egység körcikkét – figyelemfelkeltés céljából – „kihúzhatjuk" a körből. Viszont mellőzzük a kördiagram „megdöntését", mivel eltorzítja az arányokat.

A kördiagram altípusát jelenti a *kör-kör, kör-sáv* diagram, melynél egy körcikket tovább bontunk további egységekre (a *2.12. ábrán* például Kelet-Ausztriát). Használata ritka, mivel az ábra áttekinthetősége nem a legjobb. Készítésénél arra figyeljünk, hogy a továbbbontott körcikk területével nagyjából egyező területű legyen a kisebbik kör.



2.12. ábra *A népesség területi megoszlása Ausztriában (2000)*

- A *háromdimenziós kördiagram (torta)* optikailag abszolút nem semleges (azon körcikkek, melyek oldala is látszik, sokkal nagyobbnak tűnnek), így használatukat nem ajánljuk.
- Itt említhető meg az ún. *perecdiagram* is, ahol az értékeket az egyes körsávok területe (ill. ívhossza) adja meg. Elméleti előnye, hogy több adatsort is bemutathat, de hátránya, hogy az ábra optikailag nem semleges, mivel nem azonos területűek az egymás alatt-felett sorakozó körsávok (így például idősor ábrázolására tiltott). Használatát csak akkor ajánljuk, ha belecsempészhető és értelmezhető a nagyságkülönbség az ábrába (*2.14. ábrán* népességnagyság szerint követi egymást a három ország).



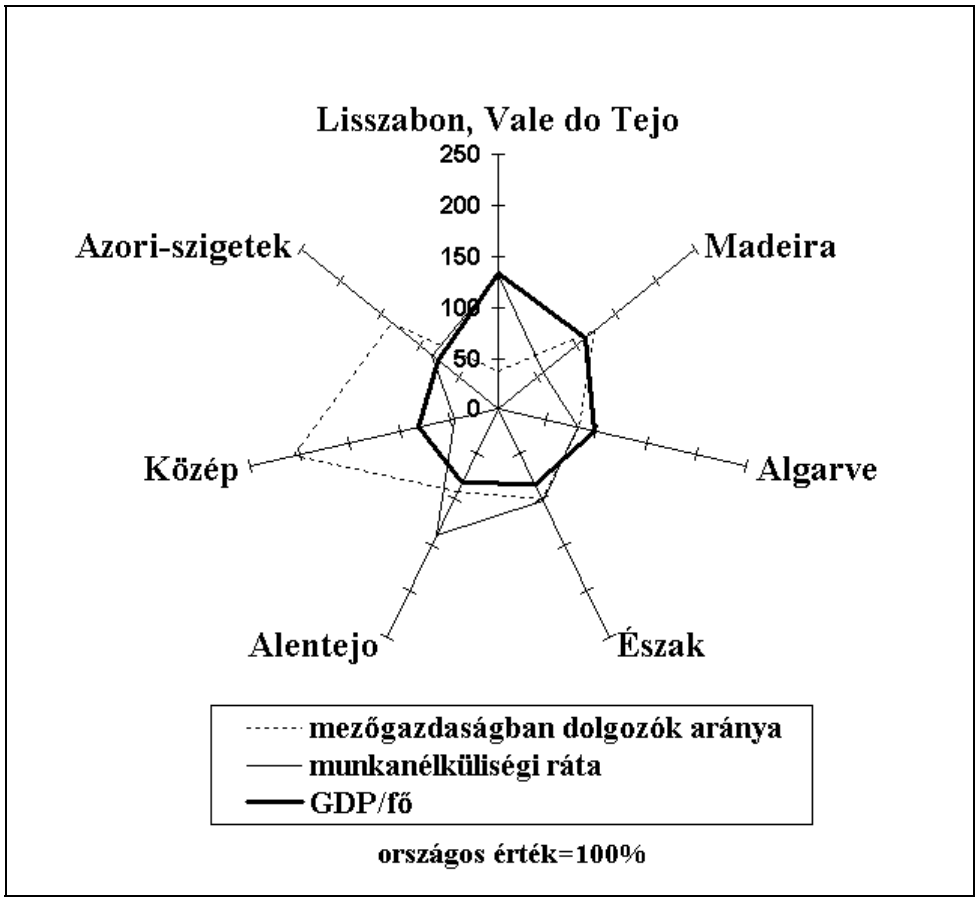
2.14. ábra *A foglalkozási szerkezet különbségei a Benelux-államokban (2000)*

2.9.4 Egyéb alakzatdiagram típusok

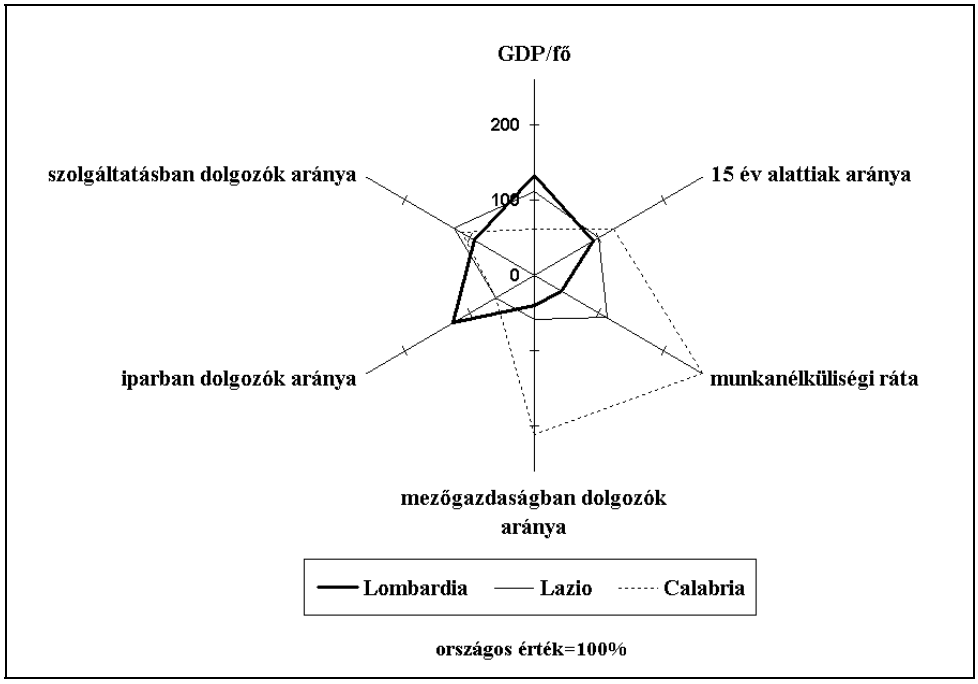
A *sugárdiagram* (esetenként *radar-diagrammnak* is nevezik) alkalmazható több területegység több (akár abszolút, akár relatív) adatsorának ábrázolására. Mivel több adatsor skálázása nagyságrendileg eltérő így, vagy azonos karakterű mutatókat (pl. foglalkozási szerkezet) ábrázoljunk, vagy hozzuk közös nevezőre a mutatókat (például az országos átlag százalékában fejezzük ki az egyes értékeket). Célszerű kevés egységet és mutatót választani, valamint az egyik fő mutató szerint sorba rendezni a területegységeket. Kétféle módon is alkalmazható:

a) Az egyes területegységeket az egy pontból kiinduló félegyenesek jelölik, míg a hozzájuk tartozó adatokat a félegyenes skáláján megjelenő pontok adják meg. Az ábra arra alkalmas, hogy a különböző mutatók egyenlőtlenségeit összevessük (*2.15. ábra*). Ezt a grafikont használja az információs társadalom egyenlőtlenségeit leíró ún. *INEXSK-módszer* → **5.8.**

b) Az egyes félegyeneseken a mutatókat ábrázoljuk, s így a területegységeket a kirajzolódó alakzatok testesítik meg. Ilyen diagramokon a hasonló illetve az elütő karakterű térségek bemutatását célozhatjuk meg (*2.16. ábra*). A legjobb helyzetű térségekhez nagyobb, a kedvezőtlen jellemzőkkel rendelkezőkhöz tartozik. kisebb területű alakzat tartozik. A minden jellemzőben nagyjából kiegyensúlyozott térségekhez tartozó poligon *konvex*, míg azokban, ahol az egyes jellemzők szerint nagyon különböző a helyzet, *konkáv* (beszögelléseket, csúcsokat tartalmazó) alakzatot kapunk.



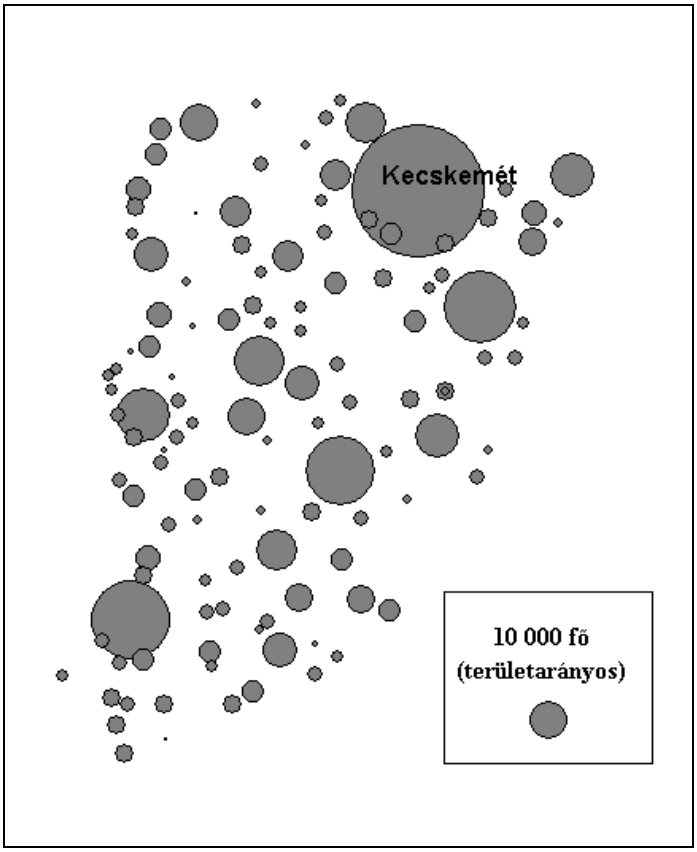
2.15. ábra Három fő indikátor regionális egyenlőtlensége Portugáliában (2000)



2.16. ábra *Lombardia, Lazio és Calabria régiók fő társadalmi-gazdasági jellemzői (2000)*

Érdekes ábrázolási lehetőségeket ad az ún. *buborékdiagram*, mely hasonlít a pontdiagramhoz, csak a pontok helyett különböző sugarú körök állíthatóak elő. A diagram három adatsor megjelenítését teszi

lehetővé, oly módon, hogy az X tengelyen az első, az Y tengelyen a második adatsor értéke olvasható le, míg a harmadik adatot a két koordináta szerinti pontba, mint középpontba megrajzolt kör területe jelenti. A területi kutatásokban alkalmazható például egyszerű tematikus térképek készítésére, ha a települések koordinátái (x = földrajzi hosszúság, y = földrajzi szélesség) és egy abszolút mutató rendelkezésünkre áll (*2.17 ábra*).



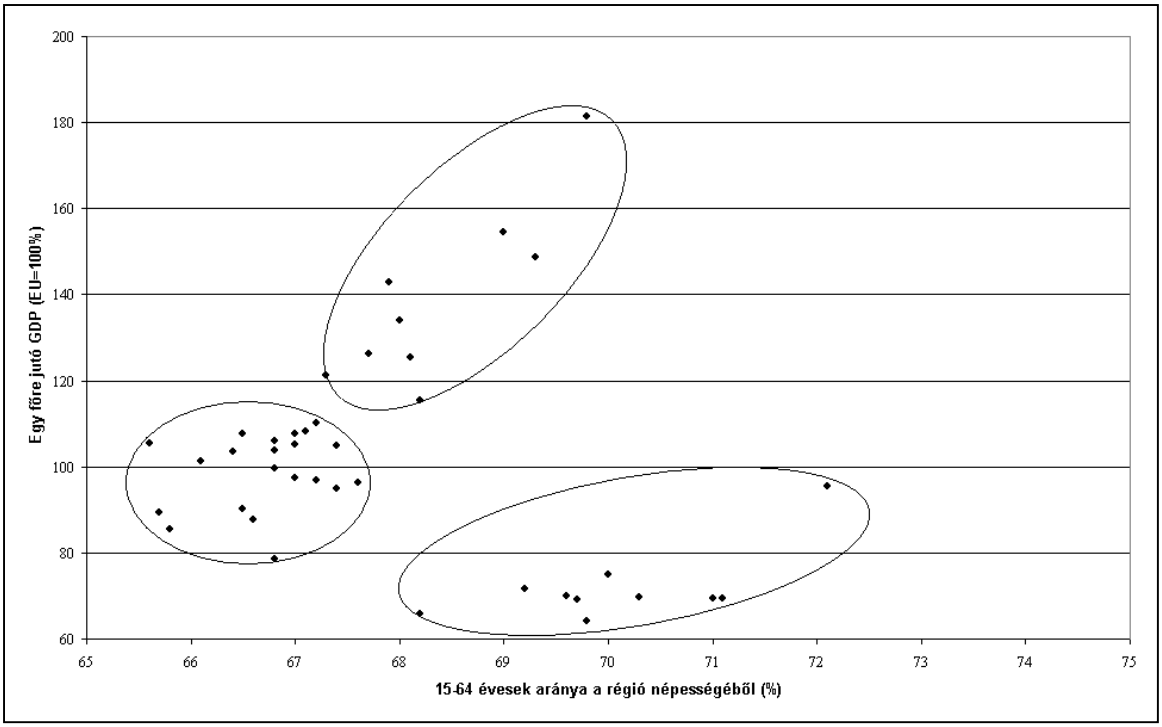
2.17. ábra *Bács-Kiskun megye népességének települések szerinti eloszlása (1999)*

2.9.5 Pontdiagramok (Jeney László)

A pontdiagram gyakran használt módszer *kétváltozós adatsorok* értékeinek grafikus ábrázolására. E diagramtípus legfontosabb alkalmazási célja: két változó közötti kapcsolat szemléltetése*, illetve* egy adatsoron belül a megfigyelési egységek csoportosítása *két változó alapján.*

A pontdiagram alapja egy derékszögű koordinátarendszer, amelynek két tengelyén (x és y) a két valószínűségi változó értékeinek nagyságát mérjük, a diagram területén lévő pontok pedig az adatsorban szereplő megfigyelési egységeket helyezi el e viszonyrendszerben. Az egymáshoz hasonló megfigyelési egységeket reprezentáló pontok a diagramon egymás közelében vannak.

A két kategóriatengely közül – amennyiben ez tartalmilag eldönthető - a vízszintesen (x) a független változót mérjük (amely balról jobbra növekszik). Ezen független változó értékeihez viszonyítjuk a másik (függő) változó értékeit, amelyek nagyságát a függőleges (y) tengelyről olvashatjuk le (*2.18. ábra*).



2.18. ábra *Az aktív korú népesség aránya és regionális fejlettség közötti összefüggés Németország NUTS2-es régióiban (2000)*

A pontdiagram esetében (szemben az oszlop- vagy kördiagrammal, illetve grafikonnal) nincs jelentősége a vizsgált adatbázis ábrázolás előtti sorba rendezésének, hiszen itt a pontok helyzetét csupán a két változó nagysága határozza meg, így a pontok függetlenek a nekik megfelelő megfigyelési egységek táblázatban szereplő sorrendjétől.

A pontdiagram igazi funkciója a kétváltozós adatsorok megjelenítése. Ez még akkor is így van, ha elvileg az Excel képes egyetlen változó kijelölése után is pontdiagramot szerkeszteni. Egy változó esetén a vízszintes (x) tengelyre az adatsor elemeinek a táblázatban szereplő sorrendje kerül (mint független változó). A diagram tehát ennek függvényében ábrázolja adott y valószínűségi változó értékeit, aminek pedig nincs különösen társadalmi-gazdasági mondanivalója. Egyváltozós adatsor ábrázolásához érdemes a pontdiagram helyett egy másik diagramtípust (pl. oszlopdiagram) választani.

Több diagramtípusnál is felmerül, de talán a pontdiagram esetében van a legnagyobb veszélye a diagram alakjának nyújtásából vagy tömörítéséből származó torzításoknak. A fenti módosítások hatására mind a diagramon látható pontfelhő alakja (külső határai), mind pedig a pontalakzat belső szerkezete (pontok közötti távolságok) jelentősen módosulhatnak, ami félrevezethető. Az értéktengelyek léptékeinek egyenlőtlen megváltoztatása különböző mértékben befolyásolja a pontok közötti távolságokat, aminek folytán a diagramra rátekintve egyes megfigyelési egységek látszólag hasonlóbbnak vagy eltérőbbnek tűnhetnek. A legtöbb diagramtípushoz hasonlóan a pontdiagram esetében is módosítható a pontok eloszlása a tengelyek *skálabeosztásának* beállításával.

Az eddigiekben egy adatsorhoz tartozó *két változó* értékeire készített pontdiagramokról volt szó. Az Excelben lehetőség nyílik egyszerre *több adatsor* egy pontdiagramon történő ábrázolására. Ilyen diagramot csak úgy célszerű szerkeszteni, ha minden adatsor esetében a kategóriatengelyek azonos változókat jelölnek. Erre már csak azért is szükség van, mert a program szerint egy értéktengelyhez csak egy címet adhatunk meg. Több adatsor esetén nem elég, hogy azonos változóink legyenek a kategóriatengelyeken, hanem az egyes adatsorok értékeinek is megközelítőleg azonosnak kell lenniük. Ellenkező esetben egyes adatsorok pontalakzatai túlzottan összeszűkülhetnek, így nehezen értelmezhetővé válnak. Több adatsort egyébiránt csak indokolt esetben ábrázoljunk egyetlen

pontdiagramon, az egymással érintkező (vagy egymást átfedő) pontalakzatok miatt az ábra könnyen áttekinthetetlenné válhat.

2.9.6 Vonaldiagram (Jeney László)

A grafikon szó gyakran tágabb értelemben használatos. A széles közvélemény egy része ugyanis a grafikon szó jelentését kiterjeszti valamennyi derékszögű koordinátarendszer segítségével történő grafikus ábrázolásra (pont-, vonal-, oszlop- vagy sávdiagram). Itt (az Excel táblázatkezelő programnak megfelelően) *a grafikon kifejezés alatt a vonaldiagramokat értjük.*

A pontdiagramokhoz hasonlóan a grafikonokat is koordinátarendszerekben ábrázoljuk, a különbség azonban az, hogy itt összekötjük a szomszédos pontokat. Ez pedig azt feltételezi, hogy az adatsor egyes pontjai (a megfigyelési egységek) egy sort alkotva láncszerűen követik egymást, ahol a szomszédos láncszemek összeköthetők. A grafikonok tipikus ábrázolási eszközei az idősoroknak, ahol a független változó maga az idő (t), és ez kerül a vízszintes (x) tengelyre, amihez hozzárendeljük adott y valószínűségi változó (adatsor) értékeit. Az időn kívül természetesen más mutató is (pl. távolság autópálya mentén vagy folyamkilométer) lehet a grafikon független változója.

Mivel a pontok összekötésével az adatsorból (mintából) kimaradó x értékekre is igyekszünk y értékeket hozzárendelni, ezzel feltételezzük, hogy bármely pontban meghatározható az y változó értéke az adatsorban szereplő két legközelebbi érték összekötésével. Ez pedig csakis a valós folyamatok alapos ismeretében képzelhető el. Hosszú távú, azonban kevés elemszámú idősoroknál előfordulhat, hogy a pontok összekötésével akár egy évtizeden keresztül stagnálást sugallunk, figyelmen kívül hagyva az egyébként időközben bekövetkezetett komoly valódi változásokat (pl. halálozások száma esetében a második világháborút közrefogó két népszámlálás – 1941-es és 1949-es – adatai között nyilvánvalóan eltűnnek a köztes háborús évek magasabb értékei). *Túlságosan rövid idősorokra egyébként nem érdemes grafikont rajzolni!*

Egy grafikonon több adatsor is ábrázolható. Ez esetben figyelni kell arra, hogy az egyes adatsorok azonos mértékegységűek legyenek, sőt az egyes adatsorok között ne legyenek jelentős értékbeli különbségek. Több, egyazon területi egységre vonatkozó idősor esetében például a bázis- és a láncindex egymástól jelentősen eltérő mértékegységű és volumenű adatsorok egy grafikonon történő megjelenítésére és elemzésére nyújt lehetőséget. (Ekkor az ábra bal- és jobb oldalán *eltérő beosztású, mértékegységű y* tengely szerepelhet.). Különböző mértékegységű adatsoroknál célszerű az ezekből számított származtatott (pl. standardizált) adatokat ábrázolni grafikonon.

Gyakori, hogy az ábrázolni kívánt adatsorok hiányosak vagy az, hogy a független (x) változó értékei nem egységnyire követik egymást (ez például a nem pontosan azonos időszakonként ismétlődő népszámlálások miatt az idősoroknál sokszor előfordul). Ilyenkor két út van:

- vagy a *grafikon* digaramtípust választjuk, de ekkor a hiányzó éveket is szerepeltetni kell az x adatsorban (s a hozzájuk tartozó y cellába *nem 0 kerül, hanem a cella üresen marad!*), mivel ezeket elhagyva torzulnának az időtávok, ugyanolyan távolságra kerülnének a grafikonon eltérő hosszúságú időszakok
- választhatjuk az Excel DiagramVarázslóján belül a *pontdiagramot* is, ahol a pontok a vízszintes (x) tengely mentén mérve nem feltétlenül egységnyi távolságra helyezkednek el, mert ez a független (x) változó értékétől függ.

2.9.7 Háromszögdiagram (Jeney László)

Három dimenziós adatrendszerek ábrázolására már térbeli koordinátarendszerre van szükség, s ott bizony már gyakran nehezen értelmezhető a papíron, két dimenziós ábraként kirajzolódó alakzat, magasabb dimenziószámok esetében pedig egyáltalában nincs mód a hagyományos grafikák elkészítésére.

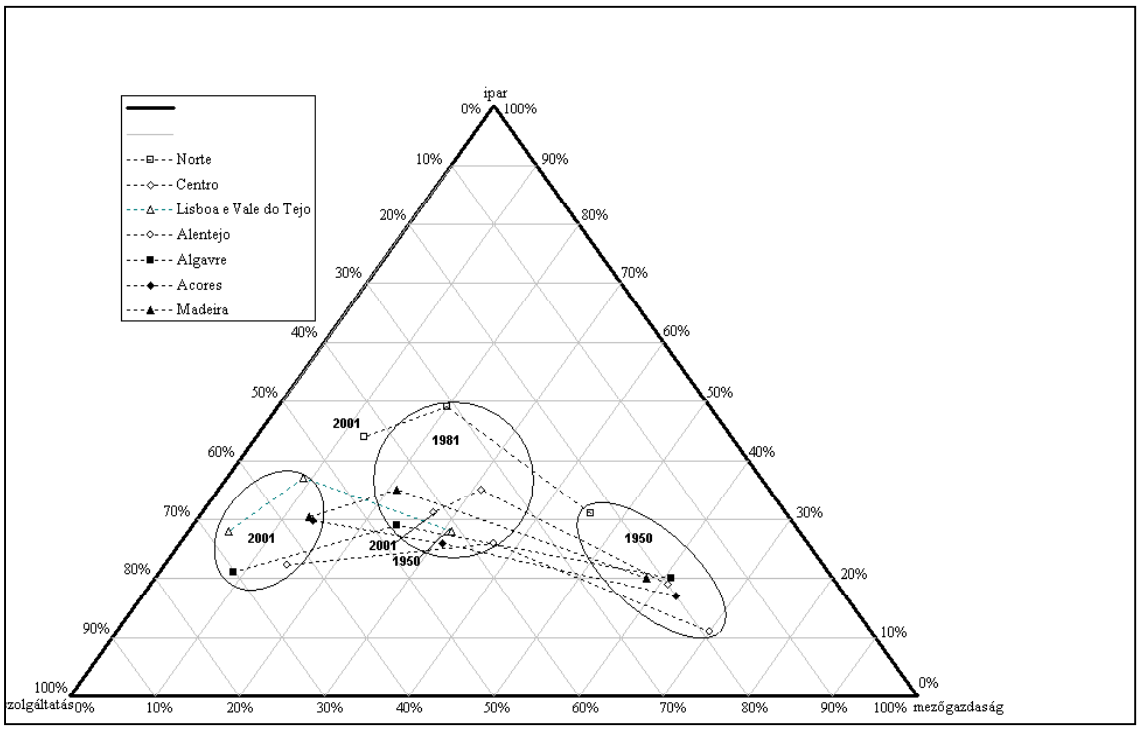
Sajátos lehetőség van olyan háromdimenziós adatrendszerek grafikus ábrázolására, ahol a megfigyelési egységekhez 3 homogén tartalmú, diszjunkt – egymást nem átfedő - volumenadat tartozik (ennek legismertebb példája a foglalkoztatottak száma vagy a termelési érték volumene a három fő szektorban – mezőgazdaság, ipar, szolgáltatások, vagy a korszerkezeti összetétel – gyerekek, aktívak, öregek száma). Az ilyen adatokból minden megfigyelési egységre kiszámítható azok megoszlása (a három komponens százalékos aránya az összegükből).

Ezek az adatok az ún. háromszögdiagramon ábrázolhatók.

Bizonyítsa be, hogy egy szabályos háromszög tetszőleges belső pontjából, párhuzamos szakaszokat húzva az oldalakhoz, az adott oldaltól az óramutató járásával megegyező irányban lévő oldallal, az így kapott szakaszok összege független a pontok helyzetétől, minden pont esetében azonos, mégpedig épp megegyezik a háromszög oldalhosszával! Ha megbirkózott a feladattal, megismerte azt a matematikai összefüggést, amire a háromszögdiagram épül.

A háromszögdiagram a legelterjedtebb táblázatkezelő programok grafikus moduljaiban nem szerepel. (Az ismertebb programok közül a STATISTICA legújabb verziója azonban tartalmaz háromszögdiagram rajzoló utasítássort). Kis ügyeskedéssel – az alapadatok átalakításával – azonban például EXCEL-ben is rajzolható háromszögdiagram (*2.19. ábra*).

A háromszögdiagram rokon vonásokat mutat a pontdiagrammal, az ábrázolt adatsoraink itt is pontokként jelennek meg. Alapvető különbség azonban a pontdiagramokhoz képest, hogy itt derékszögű koordinátarendszer helyett egy egyenlő oldalú háromszög belsejébe helyezzük el a pontokat. A háromszög minden oldala a három változó valamelyikének felel meg. A háromszög oldalai mentén az óramutató járásával ellentétesen végighaladva, két csúcs között 0-tól 100 százalékig növekszik az adott oldalhoz tartozó változó százalékban kifejezett értéke. Így a háromszög mindhárom csúcsa olyan pont, ahol a három változó közül az egyik 100 százalékot vesz fel, míg a másik két változó értéke nulla százalék. Ily módon nem csak a háromszög oldalai, hanem a csúcsai is elnevezhetők egy-egy változóról. Minél közelebb van egy pont a háromszög valamely csúcsához, annál nagyobb az adott csúcshoz tartozó változó részesedése a megfigyelési egység esetében. Fordítva pedig: minél közelebb van egy pont a háromszög középpontjához (ahol mindhárom változó épp azonos arányt képvisel), annál kiegyenlítettebb a vizsgált három változó eloszlása az adott megfigyelési egység esetében. A háromszög oldalain elhelyezkedő pontok esetében az egyik változó 0 százalékos részesedéssel rendelkezik, a másik kettő összege pedig 100 százalék.



2.19. ábra. *A régiók foglalkozási szerkezetének változása Portugáliában (1950-2001)*

Tegyük fel, hogy, hogy a három változó „a”, „b” és „c” (a+b+c=100 százalék)! Az „a” változó értékeit a lapjára fektetett egyenlő oldalú háromszög alsó oldalán (balról jobbra haladva), a „b” változóét a háromszög jobboldalán (lentől felfelé) mérjük. A „c” változó értéke pedig a háromszög bal oldaláról olvasható le, ahol a felső csúcs értéke 0, a bal alsó pedig 100 százalék. Az Excel program DiagramVarázslója nem kínál fel külön diagramtípust háromszögdiagram készítésének céljára. A pontdiagram típust választva némi „trükkel” azonban Excelben is elkészíthető a háromszögdiagram.

Ehhez a háromszögdiagram „koordinátáit” át kell alakítani a pontdiagram számára:

csúcs	Változó			**Koordináta**	
	„a”	„b”	„c”	**„x”**	**„y”**
jobb alsó	*100 %*	*0 %*	*0 %*	***100 %***	***0 %***
felső	*0 %*	*100 %*	*0 %*	***50 %***	***86,6 %***
bal alsó	*0 %*	*0 %*	*100 %*	***0 %***	***0 %***

A háromszögben elhelyezkedő pontok helyzetének meghatározásához elég két változó értéke. Így van ez a pontdiagram derékszögű koordinátarendszerében is. Ha a pontdiagramon a pontok vízszintes koordinátája *x,* a függőleges pedig *y,* akkor: $x=a+b/2$ *és* $y=b/2*\sqrt{3}$.

Érdemes először egy külön adatsorként magukat a csúcsokat felvenni, és azokat összekötve megkapjuk a háromszög körvonalát. (Megjegyzendő, hogy a háromszög kerületét alkotó adatsorhoz nem elég három pont, ugyanis negyedik adatpontnak újra meg kell adni az először megadott csúcs koordinátáit, hogy a körvonal bezáródjon.)

A háromszögdiagram lehetőséget nyújt több területi egység egyidejű összevetésére. A hasonló karakterű területi egységeket ábrázoló pontok egymás közelébe kerülnek. Az egymáshoz közeli pontok elkülönítésével a vizsgált területi egységek csoportosíthatók. Ha egyazon területi egységnél több időpontra is rendelkezünk adatokkal, akkor a pontok időrendben összeköthetők (*2.19. ábra*), így a változás iránya alapján is jellemezhetjük megfigyelési egységeinket.

2.10 Térképek

A földrajzban a térbeliség fontosságának szembeötlő bizonyítéka, hogy szinte minden információt térképeken ábrázolnak. Az utolsó néhány évtizedtől eltekintve ezek a térképek hagyományos, *analóg* formában készültek. Előállításuk hosszadalmas, és igen nagy szakértelmet követel. Napjaink jócskán felgyorsult világában ez a papíralapú technológia már számos esetben nem tud lépést tartani a változásokkal és felhasználhatósága is korlátozott. Erre válaszul először a térképi tartalmat kezdték el számítógépekkel előállítani, mára pedig a térbeli infomációk egyre nagyobb része *digitális* térképek formájában kerül forgalomba.

A térképet, mint adatforrást nagy körültekintéssel kell kezelnünk. A térkép fizikai állapota, méretaránya, vetületi rendszere stb. mind olyan tényezők, amelyeket a feldolgozás során figyelembe kell venni, különben könnyen téves eredményekre juthatunk.

2.10.1 Geodéziai, topográfiai és földrajzi térképek (Mattányi Zsolt)

A térképek méretarány szerint három csoportra oszthatók (*Lerner J. 1992*):

- geodéziai (nagyméretarányú) térképek
- topográfiai térképek
- földrajzi térképek

Geodéziai térképek

Ezek a térképek terepi felmérések alapján készülnek. Méretarányuk 1:500 és 1:10 000 között mozog. A felmérés eredményeinek ábrázolásánál eltolást nem alkalmaznak. A térképi tartalmat minimális generalizálással állítják elő.

Topográfiai térképek

A legáltalánosabban használt térképtípus. Méretaránya 1:10 000-től 1:300 000-ig terjed. A térképi tartalom a méretarány csökkenésével egyre nagyobb mértékben generalizált, az 1:25000 méretarány alatt már a térképek nem közvetlen terepi felmérés alapján készülnek. A topográfiai térképek egyszerre többféle tematikát tartalmazhatnak (domborzat, vízrajz, infrastruktúra, felszínborítás stb.). Ezeknél a térképeknél gyakran előfordul, hogy egyes objektumokat eltolással ábrázolnak, másokat kihangsúlyoznak (valós méretüknél nagyobbnak ábrázolják).

Földrajzi térképek

A földrajzi térképek kis méretarányúak (1:300 000>). Ezeken a térképeken általában csak kevés tematika szerepel egyszerre. A generalizálás igen nagy mértékű. Jobbára csak tematikus térképek alapanyagaként használhatók fel.

Egy bonyolultabb térinformatikai rendszernél előfordulhat, hogy a bemenő adatok más-más térképműről, űr-, vagy légifelvételről származnak. A különböző forrásból származó informaciók egy rendszerbe való integrálásához (geokorrekció) a modern térinformatikai és távérzékelési alkalmazásokban megvannak a megfelelő eszközök.

2.10.2 Tematikus térképek

A térképjelek formája, részletezettsége, információtartalma a méretaránnyal változik.

A regionális elemzés legalapvetőbb eszközei közé tartoznak a különböző *tematikus térképek*. A regionális kutatásban ez leggyakrabban *térségi* adatok ábrázolását jelenti.

Bár sokaknak határozott véleménye van például a térképezés során használandó *jelkulcsokról vagy színskáláról*, ezek jó része egyéni megítélés, illetve egyéni esztétikai érzék kérdése. A színes tematikus térképeknél mindesetre jó iránymutatást ad a *hideg-meleg* színskála (kék – zöld – sárga - piros és a megfelelő átmenetek). Bár elterjedt, de nem „tudományos" törvény ellenben, hogy a „pozitív", kedvező helyzetet tükröző értékekkel rendelkező területegységek kapják a meleg (piros), míg a kedvezőtlen helyzetűek a hideg (kék) színeket (az ellentétes összefüggés is megmagyarázható). Az egyszínű jelkulcsnál a *sötét-világos* átmenet felel meg az értékkülönbségek érzékeltetésének.

A térségi alapú tematikus térképeken elsődlegesen *fajlagos, relatív adatok* ábrázolandók. *Abszolút* volumeneket (például a népességszámot vagy épp a külföldi tőke nagyságát) helytelen a területegységek színezésével ábrázolni. Erre megfelelő mód az adott abszolút ismérvértékkel *arányos szimbólumokkal* (körök, négyzetek) készült kartogram.

Bár az adathiány a modern regionális elemzések térképein már viszonylag ritka, ez üresen, „fehéren" hagyott térrészekre írt n.a. – *nincs adat* – jelöléssel érzékeltethető. Sokan épp ezen problémakör miatt általában kizárják – a fejezet szerzője ellenben előszeretettel használja - a fehér, üres jelkulcs-kategóriát a tematikus térképezésből. Főként az egyszínű térképek esetében ezzel egy jól értelmezhető jelzést hagynak veszendőbe menni az adatsor elején vagy végén álló területegységek jellemzőinek azonosítására.

A különböző társadalmi ismérvek térképezése az információk vizuális megjelenítését célozza, méghozzá *információ-tömörítéssel*.

Manapság már ritkán fordul elő (aki példát talán rá, jelezze!), de azért talán nem felesleges utalni arra, hogy szép színes térképet kaphatunk akkor, ha például egy megyei adatsort úgy térképezünk, hogy *minden megyét más-más színre festünk* (abból a „logikából" kiindulva, hogy sok adatsorban valóban minden megyéhez más-más érték tartozik). Ez azonban nem tematikus térkép, hanem színes illusztráció („nyuszitojás"). A tematikus térképezés ugyanis az említett információtömörítés nélkül, aminek az a célja, hogy a vizsgált jelenség térbeli eloszlásának jellegzetességeit kiemelje, s ezáltal utaljon *a jelenség lényeges térbeli sajátosságaira* – értelmét veszti. A tematikus térképnek nem kell vizuális formában *megismételnie* a területi adattáblát. Akit egy-egy megye pontos ismérvértéke érdekel, úgyis az eredeti adatokhoz fordul.

Két, gyakran visszatérő módszertani kérdésre mindenképp érdemes kitérnünk.

Miután a megfelelő alaptérképek rendelkezésre állnak, a térképezés általában azzal indul, hogy az adatokat osztályközökbe soroljuk. Itt két kérdés vetődik fel:

- *Hány osztályközzel, kategóriával dolgozzunk?*
- *Hogyan különítsük el a kategóriákat?*

Mindkét kérdésre különböző gyakorlatias válaszok vannak, több lehetőség közül választhatunk.

Az osztályközök (kategóriák) számát az emberi vizuális észlelés határai korlátozzák. Gyakorlati tapasztalat az, hogy *nyolcnál több kategória* megkülönböztetése már nehézkes, ezért ezt a határt – még ha ügyesen megoldható is a különböző kategóriák megkülönböztetése, például *színekkel* – érdemes kerülni. A kategóriák száma függ a térképezési alapegységektől, a megfigyelési egységek (területegységek) számának növekedésével nőhet a megkülönböztetett kategóriák száma is. A szakirodalomban a megfigyelési egységek száma (n) és a

térképezéskor használandó *kategóriák számának felső határa* (k) közötti kapcsolat meghatározására a következő függvényt javasolják (természetesen ennél kevesebb kategória is használható):

$$k = 1+3,3*\log n$$

Az összefüggés néhány esetre a *2.8. táblázatban* összefoglalt értékeket adja.

Területegység (n)	Maximális kategória-szám (k)	Kerekített k	Példa
7	3,788	4	Hazai régiók
20	5,293	5	Megyék
150	8,181	8	Kistérségek
200	8,593	9	A világ országai
257	8,952	9	Hazai városok
3200	12,566	13	Hazai települések

2.8. táblázat *A tematikus térképi kategóriák és a megfigyelési egységek száma közötti összefüggés*

A másik kérdést, azaz, hogy *miként határozzuk meg a kategóriahatárokat*, mivel ez már tulajdonképpen egy általánosabb problémakört, a *tipizálás, csoportosítás* módszerét is érinti a megfelelő fejezetben tárgyaljuk és mintapéldán ott mutatjuk be(→ **6.3**). Mindkét kérdéskörhöz kapcsolódik az is, hogy a kategorizálást az adatsorokban található értékékek is befolyásolják, nemcsak a területegységek száma (*kicsiny szórású* adatok térképezésénél kevés kategória használata indokolt).

2.10.3 Topológikus térképek

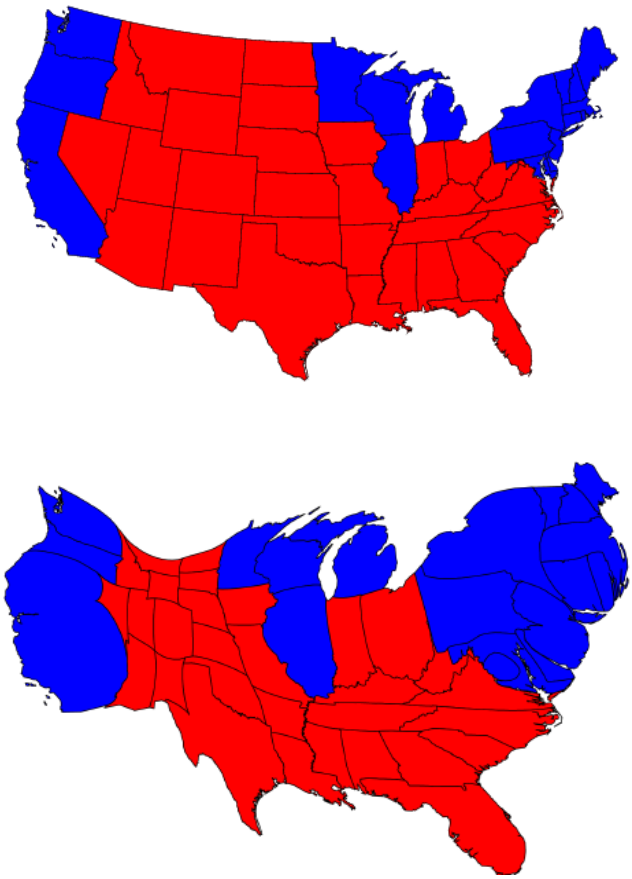
A regionális elemzések talán leggyakrabban térképezett adatai a *népességre vetített* fajlagosok. Amikor ezeket térképezzük, alapként területarányos egységekből álló (közigazgatási, regionális, országokra osztott) térképeket használunk. Ennek következtében az elkészült tematikus térképeken a legnagyobb (így sokszor a legszembetűnőbb) foltot, a nagy területű egységek jelenti, ezek uralják a térképet még akkor is, ha az ott élő népesség száma csekély. Magyarországon ez a jelenség a Dunántúl aprófalvas, az Alföld nagyhatárú településekből álló szerkezete következtében a települési szintű tematikus térképeken egyértelműen jelentkezik. Világméretekben „klasszikus" példája volt e problémának a hatalmas területű – 22mó km^2 – Szovjetunió nagy foltja bármely tematikus világtérképen.

Ennek a feszültségpontnak a feloldására alkalmasak az ún. *topológikus* (kifejezetten *nem topográfiai*!) térképek, amelyekben a „topológiát" az jelenti, hogy *az eredetileg szomszédos területegységek ezeken is szomszédosak*, ellenben azok *nagysága* nem a területtel, hanem *a népességgel* (vagy bármely más ésszerű társadalmi-gazdasági volumennel) arányos.

Ezek megszerkesztése nem könnyű, de megoldható. Az elemi megoldáskor a rajzolást úgy érdemes kezdeni, hogy a megfigyelési egységeknek kiszámítjuk a teljes területen belüli népességarányát (százalékban), majd egy száz egységnyi területet ezek szerint osztunk részekre. (Ezt a száz egységnyi alapterületet indulásként úgy érdemes megszerkeszteni, hogy alakja nagyjából kövesse a vizsgált térség – például az ország – földrajzi alakját. Gyakori megoldás az is, amikor a térkép a területegységek alakját nem őrzi meg, hanem arányos téglalapokkal vagy más idomokkal helyettesíti.) Az így kapott térképeken a belső területegységek, a határok a földrajzi, területi kiterjedésükhöz képest természetesen torzulnak, de szomszédság fennmarad. Iterációval, kísérletezéssel vizuálisan is elfogadható új térfelosztáshoz juthatunk (erre példa található: *Nemes Nagy J. szerk. 1977*).

Ha ilyen alaptérképet töltünk aztán ki a megfelelő fajlagos értékek kategóriái szerint, akkor nemcsak a fajlagos mutató szerinti jellemző (például a jövedelem, a gazdasági fejlettség szintje, vagy ahogy a

2.20. ábrán tették, választási eredmény), hanem *az érintett népesség nagysága* is érzékelhető lesz a térképeken. Nemcsak arról kapunk így információt, hogy melyik területen élnek szegények (vagy gazdagok), hanem arról is, hogy hol koncentrálódik a két pólus. A topológikus térkép helyénvaló használatára mutat példát a *2.20. ábra*, ahol a szomszédság (azaz a topológiai reláció) megmaradt, noha a területegységek nagysága megváltozott. (a fekete „peremeken" győztes Kerryt a szürke „centrumot" uraló Bush legyűrte).



2.20. ábra *A 2004. évi amerikai elnökválasztás eredményének két térképe (fenn: a szokásos területarányos, lenn az elektorarányos topológikus térkép*)

F: http://www.personal.umich.edu/mejn/election/

Ilyen logikájú topológikus térképek azonban nemcsak a fajlagos adatok ábrázolásakor kerülnek szóba. A jövedelemegyenlőtlenségek például *két topológikus térkép* egymás mellé tételével is érzékeltethetők. Az egyik lehet a *népesség számával,* a másik pedig *a jövedelem volumenével* arányos területnagyságú egységeket (régiókat) tartalmazó térkép. E két térkép összevetésekor a magas jövedelmű térségek a jövedelemtérképeken, az alacsony jövedelműek a népességi térképeken fednek le nagyobb területet. (Világméretű térképek esetében Kína a jövedelmi térképen jóval kisebbre zsugorodik Japánnál, ellenben a népességarányos topológikus térképeken természetszerűen dominálja egész Ázsiát. Szingapúr ellenben, amely a népességarányos térképeken szinte nem is látszik, nagy foltként bukkan fel a jövedelmi térképen.)

A topológikus térképek másik típusa az *elérhetőség, a közlekedés elemzése* során bukkan fel. Jól érzékeltetik a közlekedés által "torzult" teret az olyan térképek, ahol a területegységek – miközben itt is fennmaradnak az eredeti szomszédsági relációk – aszerint torzulnak, hogy milyen messze vannak egymástól az utazási időt tekintve. A sűrű kapcsolati, infrastrukturális

hálóval összekötött, központi térségek ezeken a térképeken összezsugorodnak, míg a perifériku s, nehezen elérhető térségek, megnagyobbodnak, s így távolabb kerülnek a többitől *(Gatrell, A. 1983).*

2.10.4 A mentális térkép (Kiss János Péter)

Annak a szemléletnek a korlátozott érvényét, amely *a teret* az emberek cselekedeteinek kereteként szolgáló változatlan, abszolút, objektív képződményként tekinti, a hatvanas évek közepén az Egyesült Államokban megjelent behaviorista geográfiai iskola képviselői ismerték fel először, és azóta sokan tettek kísérleteket a szubjektív terek – elsősorban a településekről, ritkábban a régiókról, az országokról a fejekben élő ún. kognitív térképek – megismerésére. (A külföldi próbálkozásokról lásd *Cséfalvay Z. 1990, 1994* összefoglaló munkáit, a hazai példák közül *Kiss J.–Bajmócy P. 1996, Bajmócy P. – Csíkos Cs. 1997*) Ez azonban nem egyszerű, hiszen a statisztikai adatgyűjtés hagyományos eszközeivel szemben itt kizárólag mintavételes eljárások, puha módszerek alkalmazása kerülhet szóba. A szubjektív terek kutatása tehát mindenképpen nagy apparátust, jelentős időbeli ráfordítást igényel, ezért nem is vált részévé a területi kutatások mindennapi gyakorlatának.

A szakirodalomban gyakran a mentális térképezés, *mentális térkép* fogalmak szinonimájaként használják a kognitív térképezés, *kognitív térkép* kifejezéseket is. Célszerűbb azonban a kettőt egyértelműen elhatárolni egymástól: a kognitív térképezés alatt a személyekben lejátszódó pszichológiai folyamatot, míg mentális térképezés alatt a kognitív térképek előhívásának módszerét értjük. A lerajzolt mentális térképek ebben az értelmezésben egyértelműen kevesebbet jelentenek tehát a kognitív térképeknél: az adott egyén kognitív térképének „felszínre hozott” részét jelentik.

A szubjektív terek vizsgálatának kulcsproblémája az ún. *kognitív térképek* – térképi elemekből illetve a hozzájuk kapcsolódó szubjektív képzetekből álló tudati képződmények – minél teljesebb "előhívása" a fejekből. A jelenleg ismert módszerek egyike sem képes e kétféle információtípust egyszerre, összefüggéseiben megragadni. Egyik részük inkább arra alkalmas, hogy feltárja: a vizsgált hely vagy régió mely részletei vannak jelen nagyobb, és melyek kisebb valószínűséggel, részletességgel, illetve pontossággal az egyes társadalmi csoportokban. Az ezt célzó eljárásokat mentális térképezésnek hívjuk, ami a gyakorlatban *kísérleti* személyekkel térképek rajzoltatását, és e rajzos információk értékelését jelenti. A módszerek másik része a vizsgált helyek (régiók) általunk előre megadott elemeihez kapcsolódó vélemények, ismeretek, attitűdök, sztereotípiák feltárását célozza. Ehhez általában speciális *kérdőíveket* használnak.

A *mentális térképek* rajzoltatása módszertani kategóriáját tekintve a *kísérletek* közé tartozik. A kísérlet a természettudományok jellemző módszertani eszköze, a társadalomtudományok közül egyedül a pszichológiában és rokon tudományaiban (szociálpszichológia, pedagógia) van jelentős szerepe. A regionális kutatásokban a mentális térképezés jószerivel az egyetlen, szélesebb körben alkalmazott kísérletes eljárás, és ezt is a pszichológiából vette át első alkalmazója, *Kevin Lynch*, illetve az ő nyomán későbbi követői a behaviorista geográfusoktól a régió- és helymarketinggel, vagy a turizmussal foglalkozó gyakorlati szakemberekig. A módszer sikerét egyszerűségének is köszönheti. Az előre kiválasztott – az általunk megismerni kívánt célcsoporto(ka)t reprezentáló – kísérleti személyek azt a feladatot kapják, hogy adott időn belül egy üres papírra vázolják föl emlékezetből a vizsgálat tárgyát képező területi egység térképét, és jelöljék be, illetve nevezzék meg azokat az objektumokat, amelyek épp „eszükbe jutnak” (vagy azokat, amelyeket legfontosabbnak tartanak). A feladat tehát igen egyszerű, és az adatfelvétel kedvező feltételek esetén (sok ráérő ember egyidejű jelenléte olyan helyen, ahol leülni és rajzolni lehet) gyorsan, olcsón végrehajtható.

A mentális térképek zömét városokról, esetleg csak városrészekről rajzoltatják, de nincs elvi korlátja annak, hogy országokról, vagy akár országcsoportokról, kontinensekről készíttessünk térképvázlatot. Szintén előnye a módszernek, hogy a kutatás céljának megfelelően a kísérleti utasítás speciális megkötéseket is tartalmazhat. Ilyen lehet az, hogy a kísérleti személyek csak azokat a helyeket jelöljék

be, ahol már jártak, vagy ahol gyakran megfordulnak; esetleg azokat, ahova szívesen elmennének nyaralni; vagy ahol rokonaik, ismerőseik élnek stb. Szűkíthető a vizsgálat azzal is, hogy korlátozzuk a bejelölhető objektumok típusát (pl. csak az utakra, vagy csak jellemző épületekre, vagy csak településekre vagyunk kíváncsiak). De van lehetőség a bővítésre is (pl. kérhetjük, hogy a bejelölt települések mellé írják oda, hogy hány lakosúnak gondolják, vagy értékeljék őket egy számmal aszerint, hogy milyen gazdagnak, szépnek stb. ítélik).

Mivel a rajzokon látható mentális térképek jellemzőit a kísérleti alanyok személyiségén kívül társadalmi és földrajzi meghatározottságaik is alapvetően befolyásolják, lehetőleg a rajzos vizsgálatot is ki kell egészíteni (a rajzolásra adott idő után) egy rövid, a kitöltőre vonatkozó alapvető háttérváltozók rögzítését célzó kérdőívvel. Legalább a nem, életkor, iskolai végzettség, illetve a jelenlegi és az életút során eddig jellemző egyéb lakóhelyek rögzítése célszerű, de a vizsgálat céljától függően egyéb tényezők rögzítése is fontos lehet (a társadalmi státusz indikátorai, pl. foglalkozásra, jövedelemre utaló adatok, a kulturális státusz, az információszerzés mennyisége és főbb forrásai, a földrajzi mobilitás mértéke stb.).

A rajzos módszer előnyei mellett számos, részben nehezen leküzdhető hátránnyal is rendelkezik. Ilyen problémát jelent, hogy a tömeges egyidejű lekérdezés általában csak speciális csoportok esetében lehetséges – nem véletlen, hogy a publikált mentális térkép-vizsgálatok jelentős részében egyetemi hallgatók vagy iskolások voltak a kísérlet résztvevői. Ezek köre azonban szűk, és a vizsgálat szempontjából legérdekesebb célcsoportokat – vállalkozók, turisták vagy az adott település/régió lakossága – jól reprezentáló mintát egyszerre „leültetni” és rajzoltatni általában nehéz, néha szinte lehetetlen. A probléma megoldását jelentheti, ha pl. a megkérdezettek lakásán történő kérdőíves vizsgálat részeként alkalmazzuk a térkép-rajzoltatást, esetleg speciális alkalmak (pl. konferenciák) használhatók ki erre.

A legnagyobb gondot azonban a rajzos információk generalizálása, és a rajzok értékelése jelenti. A térképek „minősége” nagyban függ az egyén rajzkészségétől is, a szabadkézi rajz sajátosságai miatt pedig ugyanazon térképi elemek igencsak heterogén méretben, helyzetben részletezettséggel fordulhatnak elő az ábrákon. (Egy város lakóinak városukra vonatkozó mentális térképeiből nehéz megszerkeszteni a „jellemző”, „átlagos” várostérképet, hiszen az ábrákon szereplő pontok, vonalak, megnevezések nem „adhatók össze” egyszerűen). Ráadásul gyakran nem dönthető el egyértelműen, hogy két ábra eltérései mennyiben a torzító hatások (rajzkészség, a rajzolás pillanatnyi körülményei, a kutatóval való együttműködési szándék eltérő erőssége stb.) következményei, és mennyiben fakadnak a fejekben élő kognitív térképek különbségéből. Bizonyos típusú térképi elemeket pedig egyszerűbb, gyorsabb lerajzolni, mint másokat, így egy országról készített mentális térképre például valószínűleg nagyobb eséllyel kerülnek fel pontszerű elemek és neveik (pl. települések), mint vonalak (pl. utak, vasutak), vagy pláne térségek, annak ellenére is, hogy a fejekben élő, a véleményeket és cselekvéseket meghatározó kognitív térképekben nagyobb szerepet játszanak. E problémák mértéke a kutatás megfelelő előkészítésével, az adatfeldolgozást segítő praktikus ötletekkel lényegesen csökkenthető, s az érvényes következtetések levonására alkalmas adatminőség biztosítható. Meg kell jegyezni ugyanakkor, hogy a térképek adattartalmának rögzítése és feldolgozása átlagos méretű (néhány száz fős) minták esetében is nagyon idő- és munkaigényes feladat.

A mentális térképezés kulcsát a rajzok értékelése jelenti. Ezt jelentősen megkönnyíti, ha a rajzok információtartalmának minél nagyobb részét sikerül kódolni, és kvantitatív módszerekkel is elemezhetővé tenni. Ilyen információ a megadott idő alatt az adott személy által berajzolt térképi elemek összesített száma; vagy az egyes típusok előfordulási gyakorisága (pl. pontszerű, vonalas, vagy kétdimenziós elemek, szöveges elemek stb. száma, egy várostérképen pl. az utcák, az egyes intézménytípusok stb. száma), amiből kiolvasható, hogy milyen típusú térelemek a meghatározói egy adott helynek, térségnek a vizsgált csoportban. Számszerűen értékelhető az egyes konkrét térelemek előfordulási gyakorisága is, különösen a gyakran szereplő, illetve a jellemzően hiányzó térelemtípusok vagy térrészletek elemzése lehet fontos. Ezek, különösen a kérdőívi háttérváltozókkal kombinálva, akár bonyolultabb mennyiségi módszerek alkalmazására is lehetőséget adnak, amelyekkel egyrészt az adott település/térség jellegzetességei, másrészt az egyes társadalmi csoportok jellemezhetők. Sokkal

nehezebben generalizálhatók, kódolhatók, de általában jól érzékelhetők a térelemek elhelyezkedésének és méretének valósághűsége, vagy éppen a valóságostól való jellegzetes eltérései, ami azért is fontos, mert ez utóbbiak már az egyes helyek megítélésére is utalhatnak. Általában „ránézésre" is jól elkülöníthetők a térkitöltés módja, típusa, a rajzok részletezettsége is, igaz, ezek inkább a térbeli gondolkodás működésének általános jellegzetességeire utalnak. Külön elemzés tárgyát képezhetik – ha ezt kérte a kísérleti személyektől a kutató – a térképi elemeket valamilyen szempontból minősítő, értékelő információk.

A mentális térképezés alapvető jelentősége, hogy általa a társadalmi térfolyamatok hátterében álló, más módszerekkel nem, vagy csak bizonytalanul feltárható mozzanatokról kaphatunk részletes információkat. Ezáltal új szempontokkal egészíthetők ki a regionális tudomány "hagyományos" (nem kísérleti) módszereivel nyert adatok egy egész sor, gyakorlati szempontból is nagy jelentőségű probléma esetében, mint pl. a szuburbanizáció, a földrajzi identitás a települési kapcsolatok, illetve a lakóhelyi szegregáció kutatása, valamint a várostervezés, a hely- és régiómarketing stb. terén is.

2.10.5 Térinformatikai és térképkészítő programok (Jakobi Ákos)

Kevés olyan módszer és szoftver létezik, amely olyan közel állna a területi kutatókhoz, mint a térképek és a térinformatikai eszközök. Mind az eredménymegjelenítés, mind pedig az összetett területi elemzések szempontjából hasznos és fontos eszközöknek tekinthetők, s ezért kiemelt jelentőségük van a területi szempontú vizsgálatok eszköztárában. A számítógéppel segített térképkészítés, illetve a térinformatikai rendszerek és programok felértékelődése vitathatatlan az utóbbi években, évtizedben, ami egyben azt is jelenti, hogy a területi elemzéseket végző kutatótól már egyfajta minimális elvárás az eredmények ilyen formájú megjelenítése.

A *térinformatikai* programok természetesen tovább mennek az egyszerű térképkészítési feladatokon. Bár ezek a szoftverek alapjában véve erre is képesek – sőt sokan csak ezt az oldalát használják ki a programoknak – sajátos lehetőségük a területi szemléletű információs adatbázis-elemzés, azaz a megadott kívánalmak szerinti lekérdezések végrehajtása, más szóval összetett, többtényezős kérdések megválaszolása a területi egységek szelektálása érdekében. A térinformatikai szoftverek a térelemeket és a hozzájuk tartozó információkat (attribútumokat, tulajdonságokat, adatokat) kapcsolják össze, mindezt egységes rendszerbe foglalva és a lekérdezések gyors végrehajtását biztosítva. E programok – egyéb eszközökkel kiegészítve – alapját képezik a földrajzi információs rendszereknek (GIS), illetve a nagyjából hasonló területi információs rendszereknek (*TEIR* →**1.2.3**).

Digitális térképészeti és térinformatikai eszközök és szoftverek nagy számban és változatosságban találhatók a piacon. Attól függően, hogy milyen színvonalú végeredményt kívánunk elérni, illetve milyen típusú outputra gondolunk (az egyszerű térképi vázlattól a tematikus térképábrákon át a komplett kartográfiai kiadványokig) számos programot hívhatunk segítségül. A területi elemzések során elkészített térképek, mint a megjelenítés fontos eszközei a kívánalmaktól, a felhasználói tudástól és gyakorlattól, valamint a rendelkezésre álló szoftverektől és hardverektől függően változatos formában készíthetők és készülnek el.

A raszteres térképkészítő eljárások, amelyek végső soron a képpontok egyenkénti formázására alapulnak, tulajdonképpen digitális képkezelő programokban is elvégezhetők (pl. *Adobe Photoshop*). Az így elkészített termékek nem nevezhetők térinformatikai produktumoknak, mivel a megjelenített ábra mögött nem találunk lekérdezhető információkat. Leginkább „rajzoló-programoknak" tekinthetjük ezeket, azonban a digitális térképek készítése nem lehet holmi színezgető vonalhúzogató feladat. A raszter-kezelő programok széles körű elterjedtsége és népszerűsége ennek ellenére azt sugallja, hogy az ilyen típusú térképezési eljárásokat sem szabad elvetnünk. Sokan más céllal ismerkednek meg e szoftverek kezelésével, és később a térképezési műveleteik során, korábbi tapasztalataikat felhasználva már otthonosan mozognak a programokon belül. Úgy gondolhatjuk tehát,

hogy a képkezelő programokkal készített raszteres térképek jó kezdő lépésként szolgálhatnak a bonyolultabb térképkészítő szoftverek irányába.

A tematikus térképek elkészítésének másik módja az ún. *vektoros eljárás*, amely nagy vonalakban átvezet minket a térinformatikai programok közé. E programok előnye, hogy a kívülről a programba bevitt adatbázisunkat képesek összekapcsolni a program alaptérképi adatbázisával, s ezzel a térképi és adatbeli információinkat egységes rendszerré tudják kovácsolni. Ebben az összegzett térképadat-rendszerben bármilyen bevitt adat bármilyen formában könnyen térképre vihető, könnyen módosítható akár sokszori megismétlésben is. Az így elkészített produktumok általában nyomdailag precízebbek, magas színvonalúak, ugyanakkor gyakorlati jártasságot is igényelnek az elkészítés során. A vektoros alapon működő térképező rendszerek között említhetjük meg például a *Mapinfo*, a *Geomarket* vagy az *ArcView* programokat, melyek saját formátumú fájlokkal dolgoznak, viszont az elkészített anyagok bizonyos konvertálással kivihetők a programból (*2.9. táblázat*) vagy közvetlenül a programból nyomtathatók.

Adatbevitel	*Eredmény- (térkép)kivitel*
File / Open menüben Mapinfo (.tab) Excel (.xls) dBASE (.dbf)	File / Save copy as… menüben Mapinfo (.tab)
	File / Save window as… menüben Raszteres kép formátumok (.bmp, .jpg, .tif) Windows Metafile (.wmf)
Table / Import menüben Mapinfo Interchange (.mif) AutoCAD (.dxf)	Table / Export menüben Mapinfo Interchange (.mif) AutoCAD (.dxf) dBASE (.dbf)

2.9. táblázat: *Segédlet a programok közötti átjáráshoz* (Mapinfo 5.0)

2.10.6 A természetföldrajzi térinformatika egységei és szintjei (Mattányi Zsolt)

A kvantitatív térbeli elemzési módszerek robbanásszerű elterjedésével egyidejűleg (és attól valószínűleg nem is függetlenül) megjelent a térinformatika, ami mára alkalmazott tudományágként új utakat nyitott a földrajztudomány számos ága számára. Ennek köszönhetően a természetföldrajzi elemzések során (*Kertész Á. 1997*) is hatalmas mennyiségű adatot tudunk feldolgozni, egységes rendszerben tárolni, elemezni, megjeleníteni. A térinformatikai rendszerek, adatbázis-kezelő és egyéb számítógépes elemző szoftverek a megsokszorozódott adatmennyiség és az elemzési módszerek számának rohamos növekedése miatt mára nélkülözhetetlenné váltak.

Diszkrét és folytonos jelenségek

Térbeliségük alapján megkülönböztetünk folytonos és diszkrét jelenségeket. A diszkrét jelenségek egymástól egyértelműen elhatárolhatók, mivel vagy nincs térbeli kapcsolat, vagy egyértelmű, éles határ húzható köztük. A folytonos jelenségeknél ezzel szemben folyamatos az átmenet bármely két pont között. A természetföldrajzban mind a két fajtára számos példát találunk.

- A pontszerű jelenségek nem rendelkeznek kiterjedéssel (mivel a valóságban nem is léteznek), vagy térbeli kiterjedésük a vizsgálat szempontjából elhanyagolható, lényegtelen. A pontok értelemszerűen nem fedik le hézagmentesen a teret.
- A vonalas jelenségek egy dimenziósak. Hasonlóan a pontszerű objektumokhoz, itt is találkozhatunk olyan jelenséggel, ami a valóságban több dimenziós, de a felbontás, vagy az egyszerűsítés miatt csak egy dimenziós objektumként kezelik (pl.: folyó, út, vasút), és vannak olyan vonalas jelenségek, melyek a valóságban nem léteznek (közigazgatási határok, légifolyosó, stb.). Ezek az objektumok sem fedik le a vizsgált területet.
- Az *areális* vagy területi jelenségek két dimenziósak, így foltokkal, poligonokkal ábrázolhatók. A poligonok – függően a vizsgált jelenségtől – lefedhetik az egész vizsgált területet, de nem minden esetben (települések belterületei nem fedik le az egész országot).

- A *folytonos* jelenségek esetében a térben mindenhol folyamatos átmenet van két pont között, tehát felületként értelmezhetők. Folytonos jelenségeknél példaként említhetjük a domborzatot, de a talajvíz, vagy a geológiai rétegek mélységi értékei is ilyen felületként kezelhetők.

Elemi egységek a természetföldrajzi elemzésben

A térbeli elemzések problémáját két egymástól teljesen eltérő logikával lehet megközelíteni. Az egyik módszer szerint a vizsgált terület egészét felosztjuk egyforma, szabályos részekre (cella) és ezeken belül a vizsgált jelenséget jellemezzük egy mérőszámmal. A másik lehetőség, hogy a vizsgált területen lehatároljuk a vizsgált jelenségeket és ezekhez az objektumokhoz társítunk attribútumokat. A térinformatikában az előbbi logikát a tesszelációs, míg az utóbbit a vektoros adatmodell követi. *Kiss Richárd* megfogalmazása szerint a két adatmodell közti különbség: a vektoros modellben a "hol van?", addig a raszteres modellben a "mi van ott?" kérdésekre kaphatunk könnyebben választ. A raszteres és vektoros adatbázisok egymásba átkonvertálhatók. A vektor-raszter átalakítás egyértelmű művelet, míg a raszter-vektor átalakításra több módszer is használatos (*Detrekői – Szabó 2002*).

- A *vektoros* adatmodell az adatbázisban lévő objektumokat – a térbeli jelenségek felosztásának megfelelően – pontszerűen, vonalas és poligon (3D-ben test) formában tárolja. A vektoros adatmodellben az egyes objektumokat több módon tárolják. A két legelterjedtebb vektoros adatmodell fajta a spagetti és a topológiai (pizza) modell. E két vektoros adatmodell között alapvető különbség, hogy a spagetti adatmodellben az egyes objektumok közti térbeli kapcsolatok – a topológiai modellel ellentétben – nincsenek letárolva. Ez nagy méretű adatbázisoknál jelentős mértékben növeli a szomszédsági kapcsolatokra épülő elemzések futásidejét.

- A *tesszelációs* adatmodell a térben elhelyezkedő geometriai elemeket sokszögekre bontva írja le. Megkülönböztetünk szabályos és szabálytalan tesszelációt. A szabálytalan tesszelációs adatmodell egy speciális fajtájáról a későbbiekben a domborzatmodellezésnél még lesz szó. A szabályos tesszelációs adatmodellnél három fajtáról beszélhetünk: háromszöges, négyzetes, hatszöges. Ezek azért alkalmasak mivel hézag nélkül fedik le a teret. A három fajta adatmodell közül – gyakorlati előnyei miatt – széles körben csak a négyzetes adatmodell terjedt el. A négyzethálós tesszelációt grid vagy raszteres adatmodellnek nevezik (*Detrekői – Szabó 2002*).

Jellemző természeti „layerek”

Domborzat

Természetföldrajzi elemzések esetében a legfontosabb, szinte minden elemzésnél felhasznált adatforrás a domborzatmodell. A domborzat folytonos jelenség, ezért leginkább felületekkel modellezzük. A felületeket egy speciális raszteres adatmodellből állítjuk elő. Az adatmodell előállításához kétféle mintavételezést alkalmazunk. Szekunder adatok esetében (pl. analóg térképek digitalizálása) a térképen feltüntetett szintvonalakat és a magassági értékkel jelölt pontokat visszük be az adatbázisba. Ez a módszer lehetővé teszi a terep adottságainak figyelembe vételét, hiszen így a területre jellemző morfológiát, csúcsokat, gerinceket, völgyvonalakat pontosan vissza lehet adni a modellben. Az így kapott szabálytalan elhelyezkedésű ponthalmazból állíthatjuk elő a TIN (Triangulated Irregular Network) hálót. A TIN egy szabálytalan háromszögekből álló térháló, ahol a csúcspontok magassági értékei vannak tárolva (*Detrekői – Szabó 2002*). Aktív távérzékelési módszerekkel (radar stb.) végzett adatgyűjtés esetében az egész felületről egy szabályos rácsháló pontjaira kapunk magassági értékeket. Ez praktikusan jóval nagyobb adatbázist eredményez, mivel a mintavételi módszer itt nem alkalmazkodik a domborzat sajátosságaihoz, és csak a megfelelő sűrűségű ponthálózattal garantálható a kívánt pontosság (a két adatbázismodell interpolációs eljárások segítségével egymásba konvertálható).

Morfometriai „layerek”

Számos morfometriai vizsgálat létezik. Itt a területileg értelmezhető számítások közül a legáltalánosabban használtakat soroljuk fel.

- A *kitettség* a domborzatmodellből vezethető le. A DDM rács minden egyes eleménél kiszámítható lejtésének iránya. Ez az érték értelemszerűen 0-360º között változhat. Ezt az értéket rendeljük hozzá az egyes

rácselemekhez. A kitettség térképen a rácselemek lejtésértékeit tetszőleges számú osztályba sorolhatjuk, és megjelenítéskor eszerint rendelhetünk hozzá színeket. Általános gyakorlat az égtájak szerinti osztályba sorolás.

- A *lejtés* a kitettséghez hasonlóan a domborzatmodellből számítható. Itt a rácselemek lejtését határozzuk meg º-ban, vagy %-ban. Ábrázoláskor itt is osztályokba soroljuk az egyes értékeket. Lejtéstérképek esetén általában a mezőgazdaságban és talajerózió-térképezésben használt alábbi osztályozást alkalmazzuk:

0,0 % - 5,0 %	: szántóföldi művelésre alkalmas
5,1 % - 12,0 %	: szántóföldi művelésre alkalmas, de a talajerózió veszélye fennáll
12,1 % - 17,0 %	: nagyüzemi szántóföldi növénytermesztés felső határa
17,1 % - 25,0 %	: főleg csak kisüzemi szántóföldi művelésre alkalmas
25,1 % - 40,0 %	: teraszos művelésre és erdősítésre alkalmas
40,1 % <	: csak erdőgazdálkodásra alkalmas

- Az *abszolút relief* a tengerszint (erózióbázis) feletti magasság. Ennek a kiszámítása egy elemi cellára a cella méretétől függően kétféle módon történhet. Amennyiben az elemi cella a kiinduló domborzatmodell egy cellája, akkor a helyzet egyértelmű, hiszen a cella magassági adata ismert. Ha az abszolút relief elemi cellája nagyobb a domborzatmodell elemi cellájánál, akkor már valamilyen átlagolási módszert kell alkalmaznunk.

- A *relatív relief* a vizsgált területegység legmagasabb és legalacsonyabb pontjának különbsége. Itt figyelni kell arra, hogy a megfelelő felbontást válasszuk, mert hamis következtetést vonhatunk le az eredményből. Elég csak arra utalnunk, hogy ha túl nagy egységet választunk, előfordulhat, hogy a legmagasabb pont és a legalacsonyabb nem reprezentálja megfelelően a terület domborzati viszonyait.

  Képlete: $R = M_h - M_l$

- A *felszabdaltsági index* a relatív relief és az abszolút relief hányadosa. Értéke ennek megfelelően 0 és 1 között változhat (*Kertész 1972*).

- Egy terület vízrajzi elemzéséhez első lépésként a *vízgyűjtő* területeket kell lehatárolni. A vizsgált terület vízgyűjtőkre való felosztásakor a vizsgált terület nagyságához kell igazítani az adatforrás felbontását, legyen az analóg domborzati térkép, vagy digitális domborzat modell. A vízgyűjtő területek a vízfolyásokhoz hasonlóan hierarchiába rendeződnek. Egy kis forráság vízgyűjtője esetleg csak 1-2 km^2, de a Duna, vagy befogadója a Fekete-tenger vízgyűjtője több százezer km^2. Ezekkel az elemzésekkel a másik probléma a vízgyűjtő lehatárolása, vagyis a vízválasztók kijelölése. Kézenfekvőnek látszik a domborzatmodell alapján lehatárolni a vízválasztókat – az esetek többségében ez működik is – bár a síksági területeken néha rendkívül nehéz, sőt egyes esetekben lehetetlen éles határvonalat húzni. Ezenkívül a kőzetminőség (jellemzően karsztos területeken) is módosíthatja a domborzat alapján meghúzott vízválasztó vonalat.

- A *vízhálózat sűrűségét* az egységnyi területre eső vízfolyáshosszal lehet kifejezni.

$$F_s = \frac{L}{F}$$

ahol F a terület és L a vízfolyáshossz. Az eredményt km/km^2-ben szokás megadni

- A *völgyhálózat sűrűség* a vízhálózat sűrűség kiszámításához hasonlít, azzal a különbséggel, hogy itt az egységnyi területre eső völgyvonalak hossza kerül a számlálóba, függetlenül attól, hogy van-e benne állandó vízfolyás vagy sem.

A fentiekben röviden ismertetett morfometriai vizsgálatokon kívül több más, a digitális domborzatmodellezés lehetőségeit kihasználó mutató létezik. A témához kapcsolódó angol nyelvű cikkek, tanulmányok irodalomjegyzékét *R. J. Pike (2002)* gyűjtötte össze.

Felszínborítás

A felszínborítási térképezés során a vizsgált felszín teljes felületét előre meghatározott osztályokba soroljuk. Ez a művelet tulajdonképpen egy manuális raszter-vektor átalakítás, melynek során lehatároljuk az egyes területfoltokat. A kiinduló raszteres adatbázis (többnyire űrfelvétel, vagy légi fotó) annyira bonyolult struktúrájú,

hogy bár ígéretes kísérletek már vannak a folyamat automatizálására, ezidáig még nem sikerült általánosan alkalmazható módszert kidolgozni.

A vizsgálat célja alapján megállapítjuk a kívánt részletességet. Meghatározzuk a legkisebb még térképezhető objektum méretét. Ezzel összefüggésben generalizációs szabályokat állítunk fel, amely biztosítja, hogy az adatbázisunk egységes legyen. Felszínborítási térképezéshez számos nómenklatúrát kidolgoztak világszerte az utóbbi évek során. Európában 1985-ben indították a CORINE (COoRdination of INformation on the Environment) programot, mely lehetővé tette eleinte az EU, később szinte egész Európa egységes tematikus térképezését 1:100 000-es méretarányban. A CORINE program legnagyobb adatbázisa a Land Cover (CLC100). Ebben a programban Magyarország 1993 óta vesz részt.

1. szint	2. szint	3. szint
1. MESTERSÉGES FELSZÍNEK	1.1. Lakott területek	1.1.1. Összefüggő település szerkezet
		1.1.2. Nem-összefüggő település szerkezet
	1.2. Ipari, kereskedelmi területek,közlekedési hálózat	1.2.1. Ipari vagy kereskedelmi területek
		1.2.2. Út- és vasúthálózat és csatlakozó területek
		1.2.3. Kikötők
		1.2.4. Repülőterek
	1.3. Bányák, lerakóhelyek, építési munkahelyek	1.3.1. Nyersanyag kitermelés
		1.3.2. Lerakóhelyek, meddőhányók
		1.3.3. Építési munkahelyek
	1.4. Mesterséges, nem-mezőgazdasági zöld-területek	1.4.1. Városi zöldterületek
		1.4.2. Sport-, szabadidő- és üdülő területek
2. MEZŐGAZDASÁGI TERÜLETEK	2.1. Szántóföldek	2.1.1. Nem-öntözött szántóföldek
		2.1.2. Állandóan öntözött területek
		2.1.3. Rizsföldek
	2.2. Állandó növényi kultúrák	2.2.1. Szőlők
		2.2.2. Gyümölcsösök, bogyósok
		2.2.3. Olajfa-ültetvények
	2.3. Legelők	2.3.1. Rét / legelő
	2.4. Vegyes mezőgazdasági területek	*2.4.1. Egynyári kultúrák állandó kultúrákkal vegyesen*
		2.4.2. Komplex művelési szerkezet
		2.4.3. Elsődlegesen mezőgazdasági területek jelentős természetes növényzettel
		2.4.4. Mezőgazdasági-erdészeti területek
3. ERDŐK ÉS TERMÉSZET-KÖZELI TERÜLETEK	3.1. Erdők	3.1.1. Lomblevelű erdők
		3.1.2. Tűlevelű erdők
		3.1.3. Vegyes erdők
	3.2. Cserjés és/vagy lágyszárú növényzet	3.2.1. Természetes gyepek, természetközeli rétek
		3.2.2. Hangafüves, harasztos területek
		3.2.3. Keménylombú mediterrán növényzet
		3.2.4. Átmeneti erdős-cserjés területek
	3.3. Növényzet nélküli, vagy kevés növényzettel fedett nyílt területek	*3.3.1. Homokos tengerpartok, dűnék, homok*
		3.3.2. Csupasz sziklák
		3.3.3. Ritkás növényzet
		3.3.4. Leégett területek
		3.3.5. Gleccserek, örök hó
4. VIZENYŐS TERÜLETEK	4.1. Szárazföldi vizenyős területek	4.1.1. Szárazföldi mocsarak
		4.1.2. Tőzeglápok
	4.2. Tengermelléki vizenyős területek	*4.2.1. Tengermelléki mocsarak*
		4.2.2. Sólepárlók
		4.2.3. Ár-apály által érintett területek
5. VÍZEK	5.1. Kontinentális vizek	5.1.1. Folyóvizek, vízi utak
		5.1.2. Állóvizek
	5.2. Tengeri vízfelületek	*5.2.1. Tengerparti lagúnák*
		5.2.2. Folyótorkolatok
		5.2.3. Tenger és óceán

2.10. táblázat *A CLC100 felszínborítási osztályai* (a hazánkban nem alkalmazott kategóriák dőlt betűvel, FÖMI)

A CLC100 adatbázis kontinentális léptékben megfelelő felbontású és még kezelhető méretű, de kisebb területek vizsgálatára csak korlátozottan alkalmas. Ezért egy 1995–1997 között elkészült kísérleti felszínborítási térképezés után megindult Magyarországon a CLC50 program. Az ebben alkalmazott 1:50.000-es méretarány nem csak jobb geometriai felbontást biztosított, hanem a standard kategóriák részletezésével (5 szintű nómenklatúra) tematikus tartalma is bővült.

Űrfelvétel jellemzők:	
Típus	SPOT-4 XI és SPOT-4 M (egyidejűleg készült felvételek)
az interpretációban elsősorban használt spektrum sávok	látható (XI2), közeli infravörös (XI3) és középső infravörös (XI4)
Pixel méret (felbontás)	20 méter (XI), 10 méter M felvétel (vörös sáv)
felvételi időpont	1998-1999 (2000)
űrfelvétel előfeldogozás geometriai pontossága	<10 méter, Digitális Magassági Modell alkalmazásával
speciális feldolgozás	szín információ (XI) és nagy geometriai felbontás (M) egyesítése
térképi vetület	EOV (GK50 szelvényezés szerinti felbontásban)
Adatbázis jellemzők	
területi felbontás	4 hektár (állóvizekre 1 ha)
lineáris elem felbontás	50 méter
tematikus tartalom	az EU/PHARE által a 4.szinten továbbfejlesztett nómenklatúra, a hazai adottságokra adaptálva, néhány 5. szintű osztállyal bővítve
osztályok száma	mintegy 80
poligonok száma Magyarország területén	>100 000 (becslés a kísérleti projekt alapján)
geometriai pontosság	hiba: <20 méter (RMS)
tematikus pontosság	megbízhatóság: >90%

2.11. táblázat *A CLC50 program technikai paraméterei (FÖMI)*

Egy kiválasztott mintaterületen végzett összehasonlítás alapján egyértelművé vált a két adatbázis közti jelentős minőségi különbség, ami a tematikus tartalmat illeti. Ez más okok mellett a tematikus mélység és a geometriai felbontás növelésének köszönhető. A CLC100 és a hazai CLC50 program várható ismétlése folytán (a CLC100 folytatása már elindult) változásvizsgálatokra is lehetőség nyílik a közeljövőben, ami további információkat nyújt a regionális kutatásokhoz.

Az egyéb természeti tematikák közül – itt csak felsorolásként – megemlítendő: a talaj, a talajvíz, az erózióveszélyes területek léjere, a klíma, a potenciális növénytakaró, az ökológiai rendszer valamint egyéb természet- és környezetvédelmi információk.

2.10.7 A természeti és társadalmi tartalom összekapcsolt elemzése

A természeti és társadalmi jelenségek és folyamatok egymásba kapcsolódnak. Mivel tartalmukban, hordozóikban alapvető eltérések vannak, az együttes elemzés nem könnyű.

A természeti és a társadalmi tényezők közötti kapcsolatot jól példázza az, hogy a legmodernebb társadalmi térkategória, *a régió* esetében nemcsak a fogalom értelmezésében bukkan fel minduntalan (legalábbis a földrajzi szempontú közelítésben) a táji, környezeti egyveretűség, mint követelmény, hanem ez az összetevő kétségkívül fontos eleme a régióképződés egyik alapfeltételének tekinthető *térségi kohéziónak* is.

Igaz ugyanakkor az is, hogy a természeti világban és a társadalomban a *térszerveződés, a térfelosztás* is alapjaiban különbözik. A mezo- és makroszintű természeti vizsgálatok térfelosztási szemléletében a táj-hierarchia és a tájlehatárolás, míg a társadalomkutatásban a regionális (legtöbbször igazgatási, politikai és gazdasági) tagolódás a fő szempont. A két szféra együttes elemzésében ugyanakkor épp ezek a különböző felosztások jelenthetik az összekapcsolódás egyik láncszemét[10].

[10] A hazai szakirodalomban fellelhető az ország településeinek természeti tájakhoz rendelése: *Marosi S. - Somogyi S. szerk. 1990.* A két szféra sajátos kapcsolatrendszerét vizsgálták az MTA RKK ATI kutatói is *(Csatári – Kovács szerk. 2003)*

Az összekapcsolt elemzés részeként a *társadalmi* jelenségek, jellemzők vizsgálhatók *természeti (táji) térbeli keretekben.* Ha ekkor a különböző területegységek között érzékelhető különbségek mutatkoznak, ez valószínűsíti a természeti faktorok direkt társadalmi hatását. E kapcsolatrendszerből – a kutatási eredmények tükrében – több markáns természeti dimenzió társadalomformáló szerepe kimutatható:

- Általában *a földrajzi fekvés* hatása
- Globális földi méretekben *a földrajzi övezetesség*
- A *hegyvidéki-síkvidéki* dualitása a társadalmi tevékenységeknek
- A *tengerektől, folyóvizektől* való távolság

E hatások kimutatására a területi elemzés általános eszközei alkalmasak. Az ilyen típusú kapcsolatok teszteléséhez bátran fordulhat a modern kor kutatója is, még ha fel is sejlik az eredmények kapcsán néha *a földrajzi determinizmus* (rém)képe (a világméretű fejlettségi megosztottság földrajzi – fekvési, éghajlati – háttértényezőiről lásd *Gallup, J. L. – Sachs, J. D. – Mellinger, A. 1999* tanulmányát). A válasz erre abban áll: *minden* emberi, társadalmi jelenség, jellemző, folyamat nem vezethető le a természeti determinációkból, de *számos* összefüggés, folyamat mögött igenis a természeti hatások állnak.

Itt kell megemlítenünk, hogy a térelemzés modelljei, elemzési eszközei legszorosabb rokonságot az *ökológia* (illetve, ahogy annak GIS alapú eszközrendszerét mostanság nevezik, az *ökográfia*) elemzési eszköztárával mutatják (*Perry, J. N. et al* 2002, illetve a publikációnak helyt adó *Ecography* c. folyóirat tematikus száma). Természetesen ugyancsak mind a természeti, mind a társadalmi folyamatokat összefogják a *környezetvédelemmel* kapcsolatos kutatások.

A gyakorlatiasabb momentumok közül végül mindenképp érdemes utalni arra, hogy tulajdonképp gazdaságstatisztikai „gyöngyszemként" még ma is része a térségtipizálásoknak, a *támogatott területek* kijelölésének hazánkban a termőhelyek értékének immár több mint 100 éves klasszikus mutatója a földminőséget jelző *aranykorona-érték.*