

**DEBRECENI EGYETEM  
TERMÉSZETTUDOMÁNYI ÉS TECHNOLÓGIAI KAR**

*Lóki József  
Demeter Gábor*

**GEOMATEMATIKA**

**Debrecen  
2009**

Lektorálta:  
Turjányi Sándor

Irták:

**Lóki József**: (1–11., 13–16. fejezet: elmélet és geográfiai alkalmazások, kidolgozott példák és feladatok)

**Demeter Gábor**: (1.1–1.2, 2.1, 5.3, 5.4, 8–10. 12–16. fejezet: elmélet és geográfiai alkalmazások, példák és feladatok, SPSS alkalmazások)

Közreműködtek:

**Buday Tamás** (1.1, 1.2, 12. fejezetek megírásában, a 3., 5., 7., 8., fejezetek példáinak kidolgozásában)

**Pénzes János** (5.4 fejezet példáinak kidolgozásában)

ISBN

Kiadta: a Debreceni Egyetem Egyetemi és Nemzeti Könyvtár  
Kossuth Egyetemi Kiadója  
Felelős Kiadó: Dr. Virágos Márta főigazgató  
Felelős szerkesztő: Bálint Ágnes  
Készült: a DE sokszorosítóüzemében, 2009-ben  
Terjedelem: ... A/5 ív

.....

## Tartalom

	<b>Előszó</b>	
<b>1.</b>	<b>Sík- és térkoordináta rendszerek. Trigonometriai függvények geográfiai alkalmazásai. Nevezetes vonalak, távolságok és felületek</b>	<b>7</b>
1.1.	Kétdimenziós koordináta rendszerek	7
1.2.	Háromszögdiagram	21
1.3.	Háromdimenziós koordináta rendszerek	26
1.4.	Trigonometria	29
1.5.	Nevezetes vonalak	31
<b>2.</b>	<b>Gömbháromszögtani tételek és alkalmazásuk. Távolság és terület mérésének lehetőségei, számítási feladatok</b>	<b>35</b>
2.1.	Gömbi geometria	35
2.2.	Gömbháromszögtani tételek	36
<b>3.</b>	<b>A differenciál- és integrálszámítás alapjai</b>	<b>43</b>
3.1.	Differenciálszámítás	43
3.2.	Integrálszámítás	48
<b>4.</b>	<b>Földi pontok koordinátáinak átszámítási lehetőségei. Vetületi transzformációk. Mérések, számítások pontosságának vizsgálata</b>	<b>53</b>
4.1.	Transzformációk	53
4.2.	Koordináta transzformációk	55
4.3.	Magasabbrendű transzformációk	56
4.4.	Hazai térképvetületek transzformációja	59
<b>5.</b>	<b>Adatok, adattípusok, adatgyűjtés (minta, mintavételezés, és szabályai). Régi mértékegységek és átszámításuk</b>	<b>67</b>
5.1.	Adatok, adattípusok	67
5.2.	Adatgyűjtés, adatnyerési eljárások, adatforrások	69
5.3.	Mintavételezés	70
5.4.	Műveletek adatokkal	71
5.5.	Mértékegységek és átszámításuk	79
<b>6.</b>	<b>Mátrixok. Mátrixműveletek és tulajdonságaik</b>	<b>81</b>
6.1.	Műveletek mátrixokkal	83
6.2.	Mátrixműveletek tulajdonságai	85
<b>7.</b>	<b>Halmazok, halmazműveletek és tulajdonságaik</b>	<b>87</b>
7.1.	A halmaz fogalma	87
7.2.	Halmazműveletek és tulajdonságaik	88
<b>8.</b>	<b>Valószínűségszámítás</b>	<b>93</b>
8.1.	A valószínűségszámítás alapjai	93
8.2.	Műveletek eseményekkel	94
8.3.	A valószínűség fogalma	95
8.4.	Geometriai valószínűség	96
8.5.	Feltételes valószínűség	97
8.6.	Valószínűségek szorzási szabálya	98
8.7.	A teljes valószínűség tétele	99
8.8.	Események függetlensége	100

<b>9.</b>	<b>Folytonos valószínűségi változók eloszlásai</b>	105
9.1.	Egyenletes eloszlás	105
9.2.	Normális eloszlás	106
9.3.	Lognormális eloszlás	109
9.4.	Exponenciális eloszlás	110
9.5.	n szabadságfokú $\chi^2$ -eloszlás	111
9.6.	Student- vagy t-eloszlás	111
9.7.	Binomiális eloszlás	112
9.8.	Hipergeometrikus eloszlás	114
9.9.	Poisson eloszlás	115
<b>10.</b>	<b>Matematikai statisztika – hipotézisvizsgálat – illeszkedésvizsgálat</b>	119
10.1.	Hipotézisvizsgálat	119
10.2.	Az egymintás t- (Student) próba	122
10.3.	A kétmintás t próba	124
10.4.	A $\chi^2$ -próba – Illeszkedésvizsgálat	126
<b>11.</b>	<b>A becslés</b>	133
11.1.	A pontbecslés	133
11.2.	Az intervallumbecslés	133
11.3.	A legkisebb négyzetek elve	134
11.4.	A középérték becslése	135
11.5.	A szórás becslése	135
11.6.	A konfidencia-intervallum fogalma	136
11.7.	Az adatok súlyozása	136
<b>12.</b>	<b>Dinamikus fizikai földrajz</b>	140
<b>13.</b>	<b>Korreláció- és regressziószámítás</b>	149
13.1.	Az SPSS szoftver főbb jellemzői	149
13.2.	Korrelációszámítás	176
13.3.	Regressziószámítás	180
13.4.	A korrelációs együttható szignifikancia vizsgálata	182
13.5.	A regresszió becslés pontossága	183
13.6.	Idősorok elemzése	183
13.7.	A trendszámítás módszerei	186
13.8.	SPSS alkalmazások	189
<b>14.</b>	<b>Faktoranalízis, főkomponens-analízis</b>	203
14.1.	Faktoranalízis	203
14.2.	Faktoranalízis az SPSS-ben	209
14.3.	Főkomponens-analízis	215
<b>15.</b>	<b>Klaszteranalízis</b>	223
15.1.	A klaszterelemzés fajtái	224
15.2.	Klaszteranalízis az SPSS-ben	230
<b>16.</b>	<b>Diszkriminancia-analízis</b>	237
16.1	Diszkriminancia-analízis az SPSS-ben	239
	Felhasznált irodalom	243
	Mellékletek	245

## Előszó

A Geomatematika tantárgy a bolognai rendszer keretében került bevezetésre, s egyike azon szerencsés lépéseknek, melyek valóban reformnak tekinthetők. A matematika oktatása korábban nem került középpontba a debreceni földrajzi és földtudományi képzés keretein belül, a matematika-földrajz, vagy fizika-földrajz szakon végzettek száma elenyésző volt. A régi képzésben a matematika önálló tárgyként nem jelent meg, míg most a geomatematika a BSc és MSc képzésben egyaránt végigvonul, rangot kölcsönözve a tárgynak, új koncepciót tükrözve. Az új felsőoktatási rendszerben a földtudomány BSc szakos hallgatóknak tanított matematika nélkülözi a földtudomány és földrajz szellemiségét, óraszámra nem elegendő arra, hogy szilárd elméleti alapokat adjon, sem arra, hogy feladattípusokat gyakoroltasson be. A földrajz BSc szakos hallgatók matematika felkészültsége változó, s még az előbb említett alapozó képzésben sem részesülnek az első évben. E jegyzet megírásával a hiánypótlás volt cél, ennek megfelelően az elméleti anyagrész alárendelt (a hagyományos tétel - bizonyítás sorrendből a bizonyítás hiányzik), a gyakorlati alkalmazásokon van a hangsúly: az információs társadalom számára nélkülözhetetlen adatnyerési és kiértékelési módokat is bemutatjuk az SPSS használatán keresztül, lehetővé téve nagy adatmennyiség gyors feldolgozását, ezzel megkönnyítve az elemző munkáját.

E jegyzet elsősorban a földrajz szakos hallgatóknak íródott, de hasznosan forgathatják a földtudomány szakosok is. Terjedelme jól tükrözi a dilemmát: kiindulva az igényekből, a BSc képzés jellegéből és a hallgatók előzetes matematikai tudásából, e jegyzet inkább széles tudásterületet ölel fel, mintsem mélyen foglalkozik egy-egy témakörrel. (Maradtak ki így is fontos területek, mint pl. a gráfelmélet). Azonban az itt összegyűjtött ismeretanyag elsajátítására egy félév még így sem elegendő. Ebből következően más tárgyak (Bevezetés a földrajzi adatbáziskezelésbe, Földrajzi helymeghatározás), illetve az MSc-képzés keretein belül is visszatérően hasznosítható e jegyzet: a benne lévő ismeretanyag közel három félévet tölt ki, noha az egyes speciális alkalmazásokat (regionális elemzési módszerek, dinamikus földrajz) csak érintőlegesen tárgyalja.

A földrajz sokszínűségéből eredően e jegyzet igyekszik nemcsak a matematika mindazon témaköreire kitérni röviden, amelyet egy földrajzos hasznosan alkalmazhat (és jó eséllyel meg is ért), hanem a különböző alkalmazási területek (társadalomföldrajz, természetföldrajz, geológia, meteorológia) is megjelennek. Kifejezetten geomatematikai tárgyú jegyzet ugyan született nemrégiben Geiger János tollából, de túlságosan magas szintű (és geológiai szemléletű) ahhoz, hogy a biztos alapokat nélkülöző BSc-képzésben részt vevő első évfolyam tananyagául szolgáljon. Más munkák (pl. a Regionális elemzési módszerek c. könyv) túl specifikusak egy alapképzésben részt vevő számára. E jegyzet számos kidolgozott

példát tár az olvasó elé, végigvezetve a számolást, illetve illusztrálva a szoftverhasználatot, talán ebből a szempontból jelent nagyobb segítséget azon hallgatók számára, akik matematikai tudása nem nyugszik stabil alapokon, vagy éppen a gyakorlatorientáltságot hiányolják.

Köszönet illeti a lektort, Turjányi Sándort, az Algebra és Számelmélet Tanszék adjunktusát a kézirat ellenőrzéséért, pontos és gyors munkájáért.

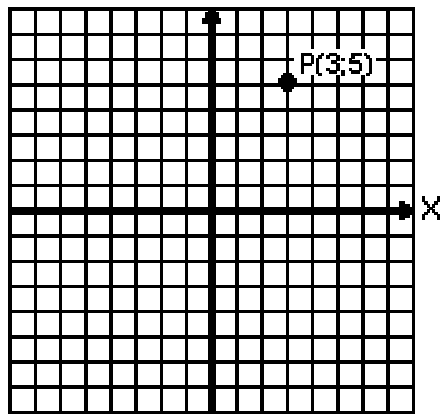
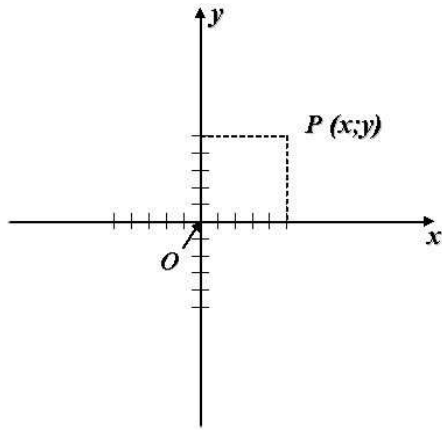
Debrecen, 2009. 04.02.

a szerzők

# 1. Sík- és térkoordináta rendszerek. Trigonometriai függvények geográfiai alkalmazásai. Nevezetes vonalak, távolságok és felületek .

## 1. 1. Kétdimenziós koordináta rendszerek

### A Descartes-féle koordináta-rendszer

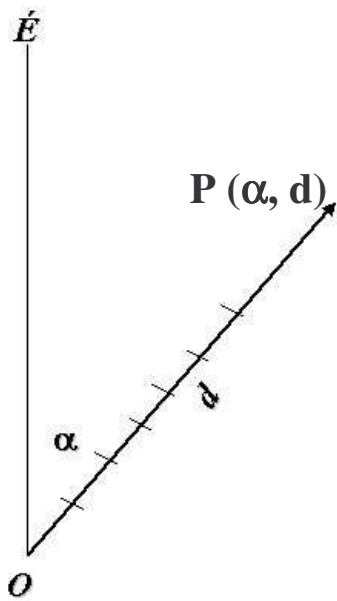


Vegyünk fel a síkon két egymásra merőleges számegyenest, amelyeknek a nulla pontja közös. Ez a pont a koordináta-rendszer origója, a számegyeneseket pedig tengelyeknek nevezzük. Az első az abszcissza tengely és a neve legyen  $x$ . A másik számegyenes az ordináta tengely és nevezzük  $y$ -nak. (A tengelyeket más betűkkel is megnevezhetjük.) A két tengely a síkot négy síknegyedre osztja. Geometriai szemléltetésben a két tengely egységét általában egyenlőnek vesszük, de ha a két tengelyre két különböző jellegű mennyiséget mérünk föl, akkor ennek nincs jelentősége. (pl. a szedimentológiában az  $x$  tengelyen logaritmikus skálát alkalmazunk)

A pontok helyét a koordináta-tengelyek síkjában a tengelyekhez viszonyított helyzete határozza meg. Vegyünk fel egy  $P$  pontot a síkon és illesszünk rá a tengelyekkel párhuzamos egyeneseket. Az  $y$  tengellyel párhuzamos egyenesnek az  $x$  tengellyel alkotott metszéspontja legyen  $x$  (a számegyenesen levő valós szám), az  $x$  tengellyel párhuzamos egyenes és az  $y$  tengely metszéspontja pedig legyen  $y$ . A  $P$  pont koordinátái  $x$  és  $y$ , amelyeket ebben a sorrendben, úgynevezett rendezett pár formájában rendelünk egy ponthoz. A sík pontjai és a rendezett valós számpárok között kölcsönösen egyértelmű kapcsolat van. Jelölés:  $P(x, y)$

A négy síknegyed pontjait egyértelműen meg lehet különböztetni a koordináták előjele alapján.

## Polár síkkoordináta-rendszer



(Erre később visszatérünk.)

Jelöljük ki egy pontot a síkon (origó), s egy kezdőirányt a kezdőpontból kiindulva (origótól). A sík egy pontjának meghatározásához ebben a koordináta rendszerben is két koordináta értékre van szükség. Az egyik koordinátát a pontnak az origótól mért távolsága ( $d$ ), a másik koordinátát a kezdő irányhoz képest elforgatott szög ( $\alpha$ ) adja.

Jelölés:  $P(\alpha, d)$

A földrajzban a rögzített irány az észak, és a szöget tőle az óramutató járásának megfelelően mérjük. (A csillagászatban a rögzített irány a dél!) A matematikában a pozitív irányú elfordulás az óramutató járásával ellentétes irányú. Tájékozódásnál a tájoló használatakor csak irányszöveget mérünk, a  $d$ -t elhagyjuk.

A polárkoordinátákból képezhetünk Descartes-féle koordinátákat szögfüggvények segítségével.

## Geográfiai alkalmazások

A **Descartes-féle koordináta-rendszert** számos területen használjuk, hiszen egyfelől térbeli jelenségek ábrázolására alkalmas, ami a földrajz számára előnyös, másrészt a két- és háromváltozós pontdiagramok (lásd excel) is ezen alapulnak. A földrajzi tájékozódásnál gyakran adódik olyan feladat, hogy egy adott, ismert koordinátájú pont helyét be kell jelölnünk topográfiai térképen, illetve a térkép segítségével meg kell határoznunk egy tereptárgy koordinátáit. A térképtanból ismert EOTR (Egységes Országos Térkép Rendszer) térképek egy olyan derékszögű koordinátarendszernek tekinthetők, amelynek kezdőpontját 200 km-re D-re, illetve 650 km-re Ny-ra tolták el a henger és a felszín érintési pontjától. Az EOVS vetületi rendszerben készült térképek vízszintes tengelye az „y”, a függőleges tengelye pedig az „x” jelet kapta. Ennek megfelelően a pont koordinátáinak általános jelölése  $P(y, x)$ . Például a  $P(800\ 000, 250\ 000)$  koordinátájú pont a kezdőponttól 800 000 m-re (800 km) K-re és 250 000 m-re (250 km) É-ra helyezkedik el.

A **polár sík koordináta-rendszert** általában akkor használjuk, ha térkép és iránytű segítségével tájékozódunk. Egy kiindulási helyről előre adott irányba,



megadott távolságra kell eljutnunk, vagy a két objektum egymáshoz viszonyított helyzetét kell meghatároznunk távolság és irány (azimut) megadásával.

**Alkalmazási területei:**

- a, vetülettan (a koordináta-rendszerek közötti szoftveres átszámításokkal külön foglalkozunk)
- b, társadalomföldrajz (ide tartoznak egyrészt a *súlypont–meghatározási* és *súlypont-eltolódási* példák, melyek egyben átvezetnek bennünket a statisztikai paraméterek alkalmazásának fontosságához is – pl. súlyozott átlagok megjelenítése a térben –, másrészt pl. *népesedési és migrációs egyenleg* vizsgálatához is használható)
- c, a talajvíz-áramlási irány meghatározása (ez átvezet bennünket a szögfüggvények használatához)
- d, geológiában a dőlés és áldőlés elkülönítése: erre a talajvíz-áramlás irányának analógiájára látunk majd példát
- e, disztancia-vizsgálatok (különbség-értékek térbeli megjelenítése)
- f, két és háromváltozós pontdiagramok (a két utóbbival az SPSS alapjainak megismerése után foglalkozunk)

**a, Vetülettani alkalmazások**

Mivel térbeli jelenségek ábrázolására alkalmas, nem véletlen, hogy pl. az EOVS és a sztereografikus koordináta rendszerek is a Descartes-féle szisztémán alapulnak (lásd részletesen a vetületi átszámítások gyakorlásánál). Más vetületi rendszerek polárkoordináta-rendszereket használnak. A szoftverek többsége x, y, z koordinátákat kér, tehát az EOVS-koordinátákat át kell váltani! A sztereografikus koordinátáknál ne feledjük, hogy a pozitív és negatív irányok felcserélődnek, tehát a „jobb felső” síknegyed rendelkezik mindkét tengelyen negatív értékekkel, míg a „bal alsó” lesz a pozitív síknegyed.

Az általunk használt *vetület.exe* nevezetű program mm élesen kéri a pontokat, ezért pl. a 396300 EOVS koordinátát 396300.000-ként kell megadni. A sztereografikus rendszerrel ne feledjük a negatív jelet kitenni, ha a papír alapú koordináta-értékeink is ezt jelzik.

**Példa:**

Egy **P** pont koordinátái Descartes-féle derékszögű koordinátarendszerben P(3;4;2). Határozzuk meg a pont koordinátáit hengerkoordinátákban, a függőleges tengelyek beosztása egyezzen meg.

Az áttérés a  $x = r \cdot \cos\varphi$ ,  $y = r \cdot \sin\varphi$ ,  $z = z$  egyenletek átrendezésével  $r = 5$ ,  $\varphi = 53,13^\circ$ ,  $z = 2$

## Feladatok

- 1, Számolja át a következő EOV koordinátákat a vetület.exe segítségével WGS84-be: 298450; 705307,  $z = 402,34$  m (azt, hogy melyik koordináta az EOV **X**, szándékosan nem adjuk meg, ugyanis ezt az értékek alapján tudni kell térképészetből).
- 2, Számolja át a következő WGS84 koordinátákat EOV-ba a vetület.exe segítségével: északi szélesség:  $47^{\circ} 34'$  és keleti hosszúság:  $20,33^{\circ}$ , magasság  $123,5$  m. (Figyelem! Az adatok két eltérő mértékegységgel vannak kifejezve, átváltás szükséges, annak megfelelően, hogy a *vetület.exe* hogyan kéri betáplálni az adatokat!)
- 3, Számolja át a következő budapesti sztereografikus koordinátákat a vetület.exe segítségével EOV-be és WGS84-be: *northing*:  $-25365$  és *easting*:  $-12875$  (a *northing* (É) és *easting* (K) angolszász kifejezést a geoinformatikai szoftverek gyakran használják).

## b, Társadalomföldrajzi alkalmazások

A *súlypontszámítás*nál a Descartes-féle koordináta rendszerrel valódi földrajzi elterjedéseket vizsgálhatunk. A súlypontszámítási módszer lényege, hogy ha a mintavételi pontokat egy síkbeli pontrendszerben (ez lehet az EOV vetületi rendszer, illetve tetszőlegesen alkotott koordináta-rendszer) elhelyezzük, a települések alappontjaihoz „harmadik dimenzióként” egy-egy „súly” (tömeg) rendelhető – mely lehet a népességszám, foglalkoztatottak száma, távbeszélő fővonalak száma, összes tőke, stb. – amelyből egy súlyozott átlag számításával megkapjuk a vizsgált jelenség súlypontjának  $x$  és  $y$  koordinátáját az adott területi egységen belül (megye, statisztikai kistérség, kistáj).

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

$x_i$  és  $y_i$  az alappontok koordinátái,  $f_i$  az alappontokhoz tartozó súly

Ez a módszer alkalmas *statikus* és *dinamikus* vizsgálatokra.

A *statikus* vizsgálatokhoz tartozik:

- egy változó súlypontjának összevetése a regionális/a geometriai központ/lakossági súlypont elhelyezkedésével,

- több változó súlypontjának összevetése a regionális központ (geometriai központ) helyével,
- több változó esetén a változók egymástól való távolsága (koncentráltság).

A *dinamikus* vizsgálatokhoz tartozik:

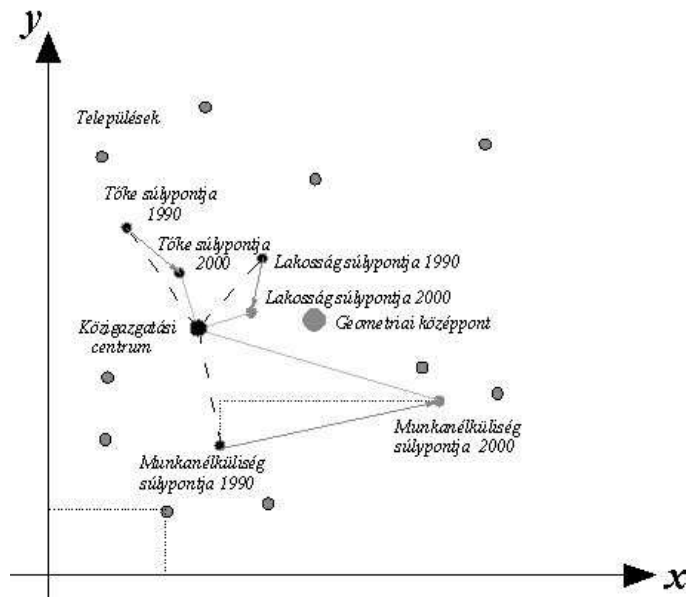
- egy változó súlypontjának mozgása az idő folyamán,
- több változó súlypontjának mozgása az idő folyamán,
- a regionális vagy geometriai centrumtól való távolság módosulásának mérése (koncentráció-dekoncentráció vizsgálata).

A vizsgált jelenség – legyen az a lélekszám, a GDP, vagy a terciér szektor súlypontja, - súlypontjának távolsága a valódi vagy ideális centrumtól (geometriai középpont) meghatározható. A vizsgálat lehet statikus és dinamikus, azaz a súlypont-eltolódás irányát és nagyságát is mérhetjük, hiszen vektorok keletkeznek. A mérés során a kezdőpont és végpont koordinátáit kivonjuk egymásból, majd a Pithagorasz-tétel felhasználásával a vektorok kiszámíthatók. Tehát AB szakasz hossza, mely A (x, y) és B(x<sub>1</sub>, y<sub>1</sub>) koordinátákkal jellemezhető:

$$AB^2 = (x_1 - x)^2 + (y_1 - y)^2$$

Egyszerre több, egymással összefüggésben lévő jelenség is ábrázolható: a munkanélküliség súlypontjának eltolódása esetén például arra számíthatunk, hogy mozgása ellentétes lesz a jövedelem, vagy a külföldi tőke, vagy akár a lakásárak súlypontjának mozgásával. A munkanélküliség súlypontjának és a lakosság szám súlypontjának együttmozgása esetén arra következtethetünk, hogy a nagy népeségszaporulattal rendelkező szegregálódó térségekben a munkanélküliség is nagy. Egy idő után viszont a gazdasági kényszerrel magyarázható migrációs folyamatok miatt megindul a lakosság elvándorlása, amelyet a munkanélküliség súlypontja ugyan követhet, de csak bizonyos késéssel.

A következő ábráról leolvasható, hogy erőteljes centralizáció játszódik le a régióban. A tőke és a lakosság súlypontja a közigazgatási centrum felé tolódott 10 év alatt, míg a munkanélküliség a centrumtól távolodva a perifériába került át. Ugyanez a súlypontszámítási módszer használható a klimatológiában (pl. a szárazsággal jellemezhető területek súlypontjának eltolódása is mérhető ezzel – ekkor a súlyfaktor a csapadék, a napsütéses órák száma stb.).



Súlyponteltolódás és a centrumoktól mért távolság

**Példa:**

A Líbiai-sivatagban négy mérőállomás mérte 1950–2000 között a csapadékokat. Ghat állomás koordinátái: K.h.10°, É.sz. 25°, a csapadék 70 mm, Ghadamesé K.h.10°, É.sz.30° és 85 mm, Murzuké K.h.14°, É.sz. 26°, 100 mm, Bengházié K.h. 20°, É.sz. 33° és 300 mm.

2000-re a csapadék rendre 50 mm, 80 mm 120 mm és 300 mm. Hová helyeződött át a területen a csapadék súlypontja? (A Föld felszínének görbületét a példában mellőzzük).

A számítás menete: súlyozott átlagot számítunk 1950-re, majd 2000-re külön-külön az x és y koordinátára. A súlyozó faktor a csapadék.

$$x_{(1950)} = (10 \cdot 70 + 10 \cdot 85 + 14 \cdot 100 + 20 \cdot 300) / (70 + 85 + 100 + 300) = 8950 / 555 = 16,1$$

$$y_{(1950)} = (25 \cdot 70 + 30 \cdot 85 + 26 \cdot 100 + 33 \cdot 300) / 555 = 30$$

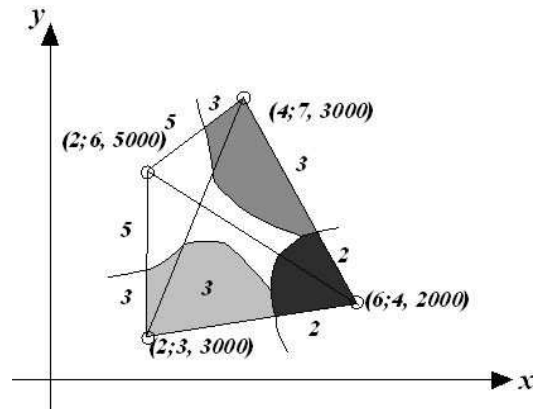
$$x_{(2000)} = (10 \cdot 50 + 10 \cdot 80 + 14 \cdot 120 + 20 \cdot 300) / (50 + 80 + 120 + 300) = 8980 / 550 = 16,3$$

$$y_{(2000)} = (25 \cdot 50 + 30 \cdot 80 + 26 \cdot 120 + 33 \cdot 300) / 550 = 30,3$$

Az elmozdulás nagysága  $v^2 = (30,3 - 30)^2 + (16,3 - 16,1)^2 = 0,13$ ,  $v = 0,36^\circ$  ( $1^\circ = 111,3$  km)

A **vonzáskörzet meghatározásának** legegyszerűbb módszerénél az előzőhöz hasonlóan járunk el. A két mintavételi pont közötti távolságot a súlyfaktorok arányában osztjuk fel, a pontos koordinátákat az előzőek alapján számoljuk. A

terület meghatározása csak azon települések esetében lehetséges, amelyek összeköttetésben vannak. (Tehát, ha **A** és **C** városok közé beékelődik **B** város, akkor **A** vonzókörzete nem terjedhet túl rajta **C** irányába). Hasonló elv alapján interpolál/extrapolál a Surfer szoftver is.



Fizikai analógiákat felhasználva e módszer alkalmas települések (v. bármely két pont) közötti vonzó/taszítóerő megmérésre. A vonzó hatás a tömeggel (lásd az előző feladatokban szereplő súlyozó tényezők, pl. lakosság, összes tőke, kórházi ágyak száma stb.) egyenesen, a távolság négyzetével (esetenként nagyobb kitevő is használható) pedig fordítottan arányos.

$$(m_1 \cdot m_2) / r^2 = G_{12}$$

Ha koordináta-rendszerben lévő pontokról van szó, a közöttük lévő távolság mérhető (pl.  $EOV_x$ ,  $EOV_y$  vagy  $\lambda$  és  $\varphi$ ). A légvonalban mért távolság helyett életszerűbb a közúton v. vasúton mért távolságokkal, még helyesebb az elérési idővel számolni, különösen ott, ahol terepakadály gyengíti a vonzó hatást (hegység, folyó, stb.).

A gravitációs modell felhasználható:

- a, települések közötti vonzóerő meghatározására (pl. A és B között nagyobb a vonzóerő, vagy a A és C, avagy B és C között),
- b, az adott pontra ható összes vonzóerő kiszámítására - ekkor az adott A pontra B, C, és D pontok által gyakorolt vonzó hatás összeadódik.
- c, ekvipotenciális vonal/felület kijelölésére - ebben az esetben azt vizsgáljuk, hogy két vagy több központi település esetében hol húzható meg a vonal, ahol gravitációs potenciáljuk egyenlő, azaz hogyan osztoznak meg a többi településen (ponton). Több település esetén vizsgálható a másodlagos szívó hatás is.

A kitevő értékének növelésével az ekvipotenciális vonal futása is változik: azt a kitevőt célszerű választani a vizsgálatban, amely már csak minimálisan változtatja meg a határokat. Az eredményeket mindig diszkutálni kell a helyes interpretáció érdekében.

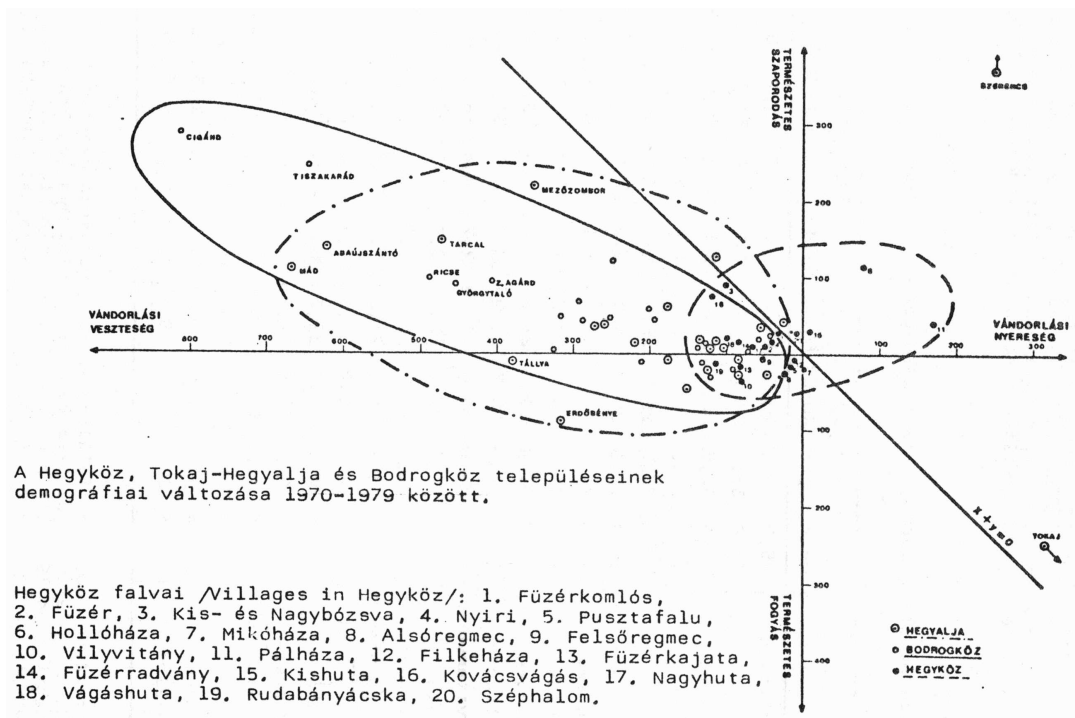
**Példa:**

„A” városban a boltok száma 100, „B” városban 200, „C” városban 300. A és B távolsága 100 km, A és C távolsága 200 km, B és C távolsága 70 km. Határozzuk meg a vonzerőt az egyes városok között és a potenciáltér nagyságát minden egyes városban ( $r^2$ -tel számolva).

$$\begin{array}{ll} G_{AB}= 100 \cdot 200/100^2= 2 & P_A= 2+0,75= 2,75 \\ G_{AC}= 100 \cdot 300/200^2= 0,75 & P_B= 2+12,2= 14,2 \\ G_{BC}= 200 \cdot 300/70^2= 12,2 & P_C=0,75+12,2= 12,95 \end{array}$$

Ha pl. B és C centrumoknak tekinthetők, akkor A pont B vonzásterébe tartozik, mert C hatása rá  $G=0,75$ , míg B-é  $G=2$ .

A *népességszám-változás és migrációs mérleg* meghatározására is használható a Descartes-féle koordináta-rendszer, amely itt fiktív térként szerepel. Az egyik tengelyen a természetes szaporodás és fogyás, a másik tengelyen a bevándorlási nyereség és veszteség szerepel, mint a szaporulatot meghatározó tényezők. Ennek köszönhetően a települések besorolhatók a létrejött négy síknegyedbe úgy, mint természetes szaporulattal és vándorlási nyereséggel rendelkező, természetes szaporulattal és vándorlási veszteséggel rendelkező, természetes fogyással és vándorlási nyereséggel rendelkező, ill. természetes fogyással és vándorlási veszteséggel bíró települések. A település lélekszámának növekedését vagy csökkenését a migráció és a természetes szaporulat alakulásának függvényében pedig az  $x + y = 0$  egyenlettel leírható egyenes jelzi, végeredményben 6 népesedési típust bemutatva. Ezt szemlélteti a következő ábra.



Települések csoportosítása népesedési jellemzők alapján koordináta-rendszerben  
(Szerk.: Süli-Zakar I.)

### Feladatok

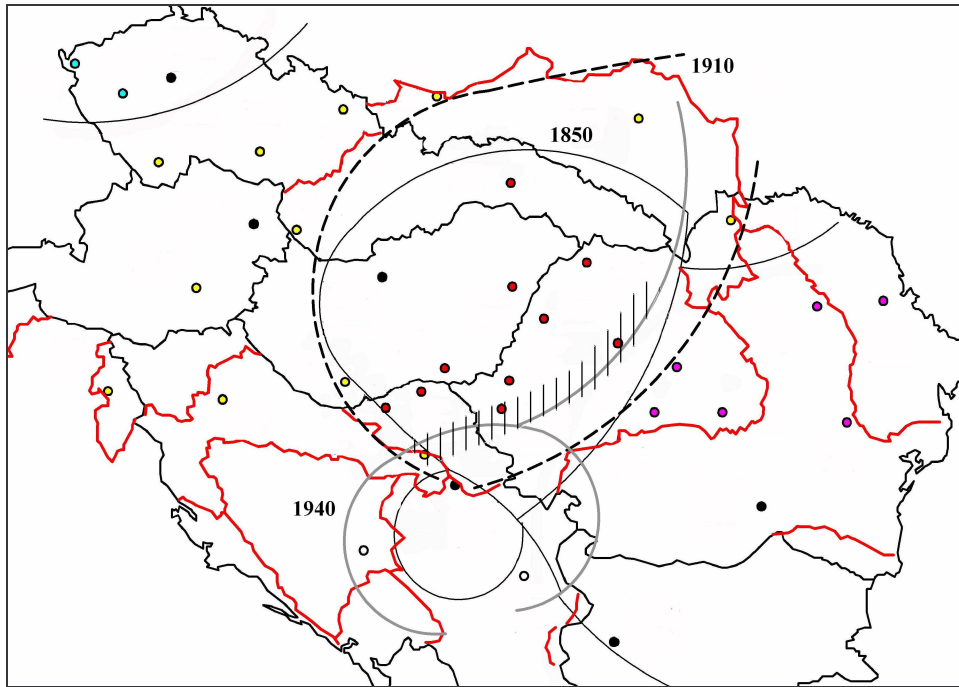
1. „A” város (koordinátái: 6, 6) ipari munkássága 10000 fő, „B” városé (3, 11) 2000 fő, „C” városé (9, 7) 1000 fő, „D” városé (0,1) 4000 fő 1990-ben. 2000-ben az adatok rendre „A”: 8000, „B”-: 4000, „C”-: 7000, „D”-: 0. Hol volt a régióban az iparban foglalkoztatottak súlypontja és hová toldott el 2000-re? Ha 2000–2010 között megduplázódik minden település ipari népessége, hová toldódik el a súlypont? Ha a 4 város lakossága először rendre 19 ezer, 23 ezer, 12 ezer és 8 ezer fő volt, majd 10 év múlva 16 ezer, 20 ezer, 14 ezer és 13 ezer, akkor az ipari népesség súlypontja közelebb került a lakosság súlypontjához, vagy sem? Mire következtethetünk ebből?
- 2, 1990-ben „A” városba (koord. 2, 2) befektetett tőke 12000 millió forint, „B” városé (3, 11) 2000, C városé (9, 4) 1000, D városé (0,2) 4000. 2000-ben az adatok rendre A: 8000, B: 4000, C: 7000, D: 0. Hol volt a régióban a tőke súlypontja és hová toldott el 2000-re? A munkanélküliség ugyanakkor rendre (ABCD): 1000, 2000, 3000, 2000 fő 1990-ben, 2000-ben pedig (ABCD sorrendben): 2000, 3000, 0, 4000 fő. Merre toldott el a munkanélküliség centruma a kistérségben? Van-e kapcsolat a munkanélküliség súlypontjának

és a tőke súlypontjának elmozdulása között? Mire következtethetünk ebből? Ha C város a közigazgatási központ, akkor a súlypontok koncentrációja v. dekontrációja zajlott le?

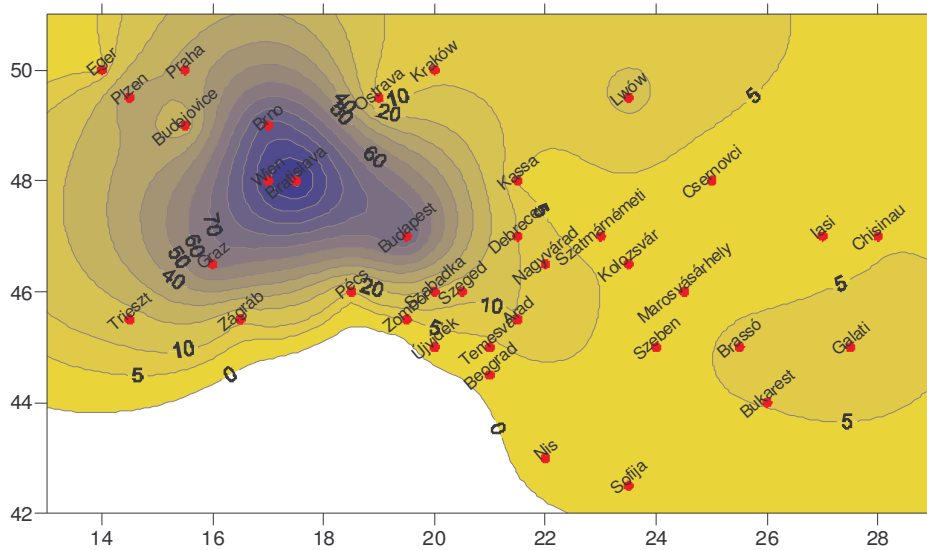
3. Helyezze el a koordináta-rendszerben a következő településeket népesedési adataik alapján:  
Halomhegy (természetes szaporulat = -300 fő, vándorlási egyenleg = +200 fő), Sárfüred (természetes szaporulat = +236, vándorlási egyenleg -256), Gyopároskedd (természetes szaporulat = +103, vándorlási egyenleg = -22)
4. „A” városban (koordinátái: 6, 6) naponta a vidékről bejáró ipari munkásság létszáma 10000 fő, „B” városé (3, 11) 2000 fő, „C” városé (9, 7) 1000 fő, „D” városé (0,1) 4000 fő 1990-ben. Határozza meg gravitációs modell alapján a települések vonzáskörzetét! 2000-ben az adatok rendre „A”: 8000, „B”: 4000, „C”: 7000, „D”: 1000. Hogyan változott a vonzáskörzet nagysága (egyenletes településsűrűséggel számolva, a települések közti távolságot a súlyfaktorok értékeinek arányában felosztva)?
5. „A” város (koordinátái: 6, 6) ipari munkássága 10000 fő, „B” városé (3, 11) 2000 fő, „C” városé (9, 7) 1000 fő, „D” városé (0,1) 4000 fő 1990-ben. Határozza meg a városok közötti (ipari) vonzerő nagyságát! Ha A és D a központi település, hogyan osztoznak meg a téren (B és C melyik vonzáskörzetébe tartozik?)
6. Helyezze el a koordináta-rendszerben a következő településeket népesedési adataik alapján:  
Kisvásárhely (természetes szaporulat éves átlaga 1990-2000 között = +3%, vándorlási egyenleg = +2%, 2000-2010 között ez rendre +1% és 0% ),  
Pénzpaták (természetes szaporulat éves átlaga 1990-2000 között = +12%, vándorlási egyenleg = -8% 2000-2010 között -6% és +1%), Virágosrét (természetes szaporulat 1990-2000 között = -5%, vándorlási egyenleg = -2%, 2000-2010 között -12% és +1%). Számolja ki, hol a legnagyobb a változás/elmozdulás az 1990-2000-es periódushoz képest!
7. Az alábbi táblázat adatai alapján határozza meg Bécs, Budapest, Belgrád és Bukarest vonzásának határait 1850-ben és 1941-ben. Számolja ki az összpontenciált a vizsgálati területen! (A táblázatot követő ábrák a megoldást mutatják be, a távolság légvonalban térképen mérve, egy egység 7 km)



lakosság (1000 fő)	1850	1910	1941	$x$	$y$	Budapest távolsága	Bukarest távolsága	Prága távolsága	Bécs távolsága	Belgrád távolsága
Trieszt	75	220	230	14,50	45,50	71,90			53,40	87,10
Zágráb	30	75	170	16,50	45,50	48,50			42,70	60,00
Graz	50	140	200	16,00	46,50	45,00		51,70	20,64	
Bécs	500	1500	1800	17,00	48,00	33,60		40,30	1,00	
Pozsony	25	75	150	17,50	48,00	23,60		47,40	10,40	
Brno	50	120	250	17,00	49,00	42,00		27,90	17,30	
Plzen	40	75	130	14,50	49,50	76,80		12,20	44,50	
Budejovice	25	60	130	15,50	49,00	62,00		24,40	27,40	
Prága	100	220	800	15,50	50,00	71,30		1,00	40,30	
Eger	10	120	120	14,00	50,00	90,30		21,54	57,90	
Ostrava	10	40	140	19,00	49,50	40,30		42,40	34,90	
Krakkó	30	150	280	20,00	50,00	45,20		64,60	53,80	
Lvov	110	250	300	23,50	49,50	72,70	97,40	114,70	96,50	
Csernovci	20	100	130	25,00	48,00	85,60	68,80		115,60	
Kolozsvár	30	60	110	23,50	46,50	59,20	52,30		92,70	
Temesvár	15	70	110	21,00	45,00	42,70	66,18		74,50	21,40
Budapest	175	930	1200	19,50	47,00	1,00	107,60	71,30	33,60	52,70
Debrecen	30	95	120	21,50	47,00	31,60	80,10		64,40	
Szeged	30	110	136	20,50	46,00	26,50	93,50		57,70	28,40
Szabadka	45	90	90	20,00	46,00	29,30	86,80		57,80	23,70
Belgrád	10	75	250	21,00	44,50	52,70	74,00		79,70	1,00
Szófia	15	100	260	23,50	42,50	103,80	48,00		133,60	53,80
Bukarest	120	400	900	26,00	44,00	107,00	1,00		140,30	74,70
Galati	35	60	85	27,50	45,00	118,70	28,80		154,50	
Iasi	40	75	120	27,00	47,00	105,60	49,80		137,90	
Chisinau	70	110	100	28,00	47,00	121,60	57,40		155,60	
Pécs	15	48	78	18,50	46,00	28,10			43,50	37,20
Újvidék	10	33	61	20,00	45,00	40,90	84,80		65,80	13,40
Zombor	10	30	35	19,50	45,50	30,80	93,70		54,20	25,00
Kassa	15	45	70	21,50	48,00	34,50			59,30	
Nagyvárad	30	61	92	22,00	46,50	39,70	69,80		73,20	
Arad	40	61	90	21,50	45,50	38,00	68,50		72,00	28,90
Brassó	32	41	60	25,50	45,00	86,80	29,30		120,00	
Szeben	20	33	53	24,00	45,00	70,70	37,00		104,00	
Nis	20	25	45	22,00	43,00	82,30	59,90		109,00	32,00



Budapest és környező fővárosok vonzásteré 1850-ben (vékony fekete), 1910-ben (vastag fekete), 1940-ben (szürke) lakosságszám alapján és kapcsolatuk a tényleges határokkal. A sraffozás az átmeneti zónát jelzi, ha  $r^2$  helyett  $r^3$ -t választunk

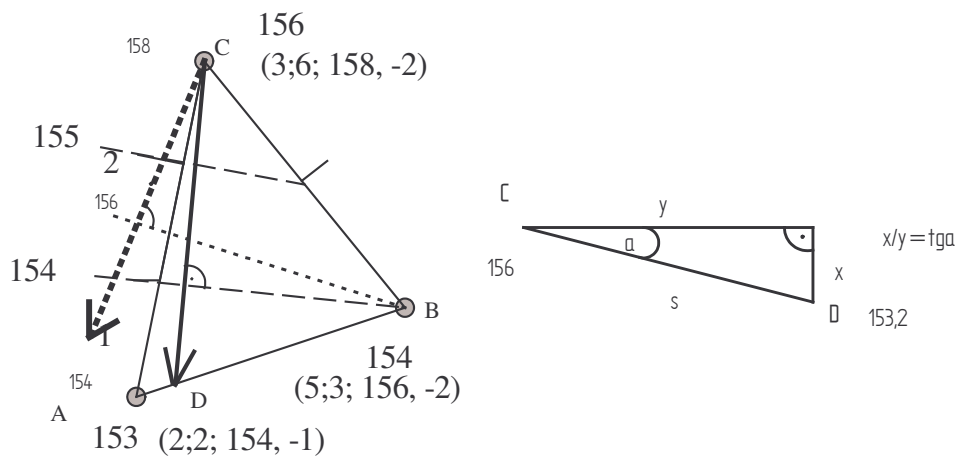


A fővárosok vonzáskörzetének határa és az összpotenciál 1850-ben koordinátarendszerben (szélességek és hosszúságok) ábrázolva

### c-d, Talajvíz-áramlási irány és a dőlésirány meghatározása

Az egyszerűbb, grafikus módszer esetében a vizsgálat elvégzéséhez legalább 3 pontra (észlelőkútra) van szükség, melyek koordinátái egy általános háromszöget határoznak meg. Ha adott a terepszint magassága és ismert a talajvízállás, akkor a talajvízszint tszf.-i magassága is kiszámítható. Ezt követően ott, ahol a legnagyobb a talajvíznívó dőlése – tehát a legmagasabb és legalacsonyabb vízállással bíró oldalon – bejelöljük a harmadik oldalhoz tartozó vízállást. Ha esetünkben a legmagasabb vízállás 156 m, a legalacsonyabb 153 m, azaz a szintkülönbség 3 m, a harmadik csúcsban pedig a vízszint 154 m, úgy a 154 m-es magasságot jelölő pont 2:1 arányban osztja fel az előző szakaszt (tehát a talajvízállás értékét súlyfaktorként használjuk). Ezt a pontot összekötjük a 154 m-es csúccsal, majd az így keletkezett szakaszra merőlegest állítunk a 156 m-es csúcsból. Ez utóbbi lesz az áramlás iránya, mely tehát merőleges az előbb meghatározott 154 m-es izovonalra (a lejtés merőleges a szintvonalakra). Az áramlás addig tart, míg a vízszintkülönbség ki nem egyenlítődik, az áramlás sebessége tehát csökken, akárcsak a nívó lejtése.

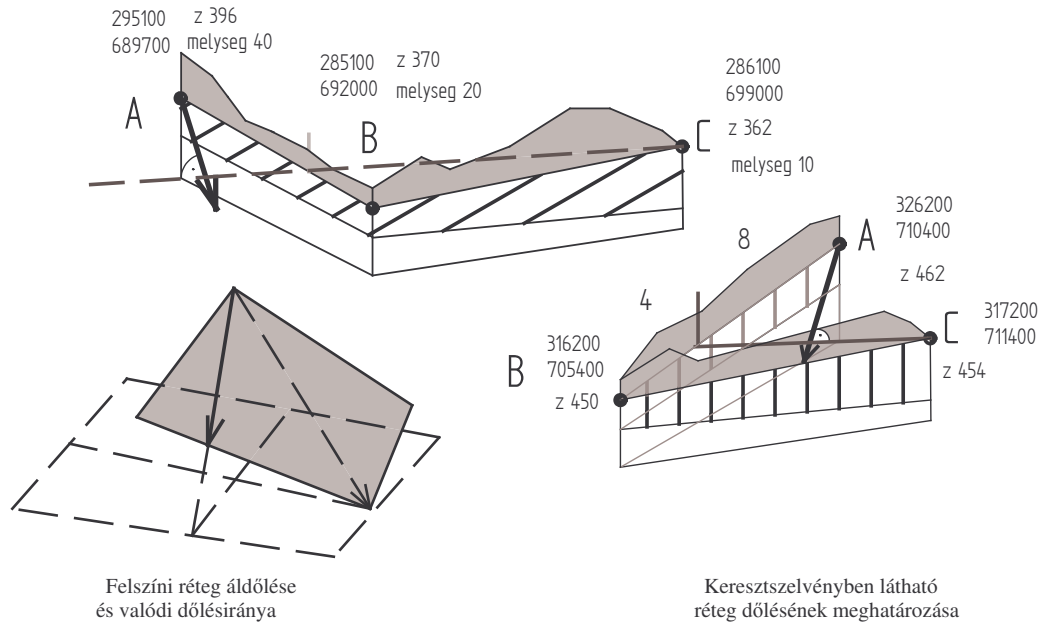
Ha ismeretesek az áramlási irányt reprezentáló vektor kezdő és végpontjának koordinátái, akkor meghatározható a megtett út (s), és az áramlás (v. szennyezés terjedésének) kezdeti sebessége:  $v = \sqrt{2 \cdot a \cdot s}$ , ahol a (gyorsulás) =  $g \cdot \sin \alpha$ . A kezdeti lejtés ( $\alpha$ ) kiszámolható, ha ismeretes a vektor kezdő- és végpontjánál pontjánál a talajvízszint tszf.-i magassága. A gyorsulás és az eltelt idő ismeretében az áramlás (szennyezés) terjedése (s) megadható:  $s = \frac{a}{2} \cdot t^2$ .



A felszín lejtésirányának (szaggatott) és a talajvíz-áramlás irányának (folytonos) meghatározása 3 észlelőkút alapján (x, y, z koordináták és a felszíntől mért talajvíz-mélység).

Előfordulhat, hogy a vektor kívül esik a háromszögön, de ez nem ad rossz eredményt. Amennyiben háromnál több kútunk van, úgy az irányt minden kút-hármasnál meg kell adni, majd a vektorokat összegezni. (Ugyanezt automatizáltan végzi el a Surfer program). Mivel nem a felszínt, hanem a vízszintet vizsgáljuk, a z koordináták értékeiből a talajvízállást mélysége kivonandó. A felszíni értékekre elvégzett hasonló vizsgálat a felszín dőlésének irányát adja meg.

A következő ábrán érintkező keresztmetszvényekben, valamint kibukkanó rétegfelszínen valódi és áldőlés látható.



A rétegfelszínek dőlésének kijelölése nehéz, hiszen a legmagasabb és legalacsonyabb pont összekötése sokszor nem a valódi, hanem az áldőlést adja meg. A dőlésirányt terepen a legegyszerűbben úgy állapíthatjuk meg, hogy vizet csepegtetünk a felületre. Amerre a víz fut, arra van a valódi dőlésirány. Ezt követően bányászkompasszal a dőlés azimutja és az irány meghatározható polárkoordináta-rendszerben.

Hasonlóan kell eljárni, ha két rétegtant bemutató keresztmetszvényünk van, melyek egy közös ponton érintik/metszik egymást és egy fedővel és fekével is rendelkező (tehát terepen nem látható) réteg valódi dőlését kívánjuk megállapítani.

Mind a talajvíz áramlási irány, mind a rétegdőlés kiszámítható vektorokkal, megadhatók a pontos koordináták is, de mivel ezt szoftverek (pl. a Surfer) is elvégzik, ezekkel itt nem foglalkozunk.

### ***Feladatok***

- 1, Egy talajvízkútban (EOV 833440, 278650, 199) a vízállás  $-2,5$  méter, egy másikban (833590, 278600, 202)  $-2,1$  méter, egy harmadikban (835110, 278940, 197) pedig  $-3$  méter. Határozza meg a talajvízáramlás irányát grafikusán, a nívó kezdeti dőlésszögét és az áramlás kezdeti sebességét.
- 2, Egy talajvízkútban (199 m) a vízállás  $-2,5$  méter, a tőle 500 méterre lévő 200m tszf. magasságú kútban  $-2,1$ m. Mennyi a talajvíz várható magassága  
a, félúton  
b, az első kúttól 350 méterre?
- 3, Egy mészköréteg tetejének EOV koordinátái egy geológiai rétegszelvényben: kezdőpont (834440, 278650, 179), végpontjái: (833590, 278600, 144). A vele érintkező másik szelvényben e réteg koordinátái a következők (835100, 278000, 205) és (833590, 278600, 144).  
Számolja ki a rétegek dőlését mindkét szelvényben, áldőlését a vízszintes síkhoz képest,  
és határozza meg a rétegek valódi dőlését és csapását (a három pont által meghatározott síkon).
- 4, Határozza meg a talajvízáramlás irányát és kezdeti áramlási sebességét a következő koordináták alapján:  
A (x,y,z, h): 1, 2, 171,  $-1$ ; B (x,y,z, h): 8, 4, 166,  $-1$ ; C (x, y, z, h): 3, 8, 170,  $-1$ !
- 5, Határozza meg a valódi dőlés irányát a következő koordináták alapján:  
A (x, y, z): 1, 2, 171, B (x, y, z): 8, 4, 169, C (x, y, z, h): 3, 8, 170!
- 6, Számolja ki a két, a fenti ábrán látható, nem feltáruló réteg valódi dőlésének szögét!

### **1. 2. Háromszögdiagram**

A háromszögdiagram a földtudományok több területén (társadalomföldrajz, szedimentológia, ásványtan, stb.) alkalmazott speciális koordináta-rendszer. Segítségével három adat ábrázolható, ha azok összege egy adott érték. Ez leggyakrabban úgy érhető el, hogy az adatokat az összegükkel elosztjuk (normáljuk, azaz százalékot képzünk) vagy eleve százalékos formában fogalmazzuk meg. Erre vonatkozó klasszikus alkalmazás a munkaerő szektorok szerinti megoszlása (mezőgazdaság + ipar + terciér és kvaternél =100%), a főbb szemcsefrakciók megoszlása (homok + iszap + agyagtartalom =100%), az ásványtanban a földpátok minősítésére használt Niggli-féle diagram, a bauxitok minősítésére használt diagram (agyagásványok, Fe-ásványok, Al-Ti-ásványok), a Selley-féle homokkőosztályzási diagram (agyag, földpát, kvarc): gyakran alkalmazzák kémiai összetétel ábrázolására is.

Nagy előnye, hogy a három adat síkban ábrázolható. Ennek oka, hogy a három adat egymástól nem független, azaz valójában két független változónk van. Használatával – csakúgy, mint a kétváltozós diagramok esetében – lehetőség nyílik

**a**, csoportalkotásra és

**b**, időbeli változások követésére.

A **csoportalkotás** során két eljárást különíthetünk el. Az egyik, hogy saját adataink elhelyezkedése alapján a vizsgált paraméterek szerint saját magunk különítünk el csoportokat, és keressük az azonos csoportba tartozás földtudományi okát. Bizonyos háromszögdiagramokban a szakirodalom által meghatározott határok különítik el az egyes kategóriákat, azaz adatainkat ezekbe elhelyezve megállapítható, hogy melyik kategóriába tartozik egy-egy adatsorunk (klasszifikációs-diagramok).

A csoportalkotásra példa az USDA textúrán alapuló talajosztályozási rendszere, a keverék törmelékes üledékes kőzeteket besoroló diagram, a Streckeisen–diagram, vagy az egyes országok fejlettség szerinti elkülönítése a foglalkoztatottak/tőke/termelt GDP szektorok közötti megoszlása esetén.

A klasszifikáció példáját először a Streckeisen-diagramon mutatjuk be. Ebben az esetben a kőzetminta adatai az alkotó ásványcsoportok %-os arányai. A diagram azonban csak három (négy) csoport tagjait tartalmazza: a kvarc, az alkáliföldpátok és plagioklászok, valamint az ún. földpátpótlók arányát. A kvarc és a földpátpótlók együtt nem fordulhatnak elő, így valójában egyszerre mindig csak három adat szerepel egy-egy minta adatainak ábrázolásában. A vizsgált kőzet ásványos alkotói közül ki kell választani az ábrázolandókat, és ezek összegét (amely a legtöbb kőzet esetében kisebb lesz a 100 %-nál) vissza kell normálni 100 %-ra. Ez azt jelenti, hogy a kijelölt alkotók százalékos értékét osztjuk az összegszázalékkal, majd szorozzuk százal, így újra százalékos érték lesz. Figyelni kell tehát arra, hogy itt a százalékokból lesznek újra százalékok! Ezután a kőzetminta összetételét reprezentáló pontot elhelyezzük a diagramba, és leolvassuk, hogy ez alapján melyik kőzettípusba tartozik a mintánk (előtte ellenőrizve, hogy az adott típusú kőzethez használhatunk-e egyáltalán ilyen módszert). A diagram érdekessége, hogy két háromszögből áll, melyeket a kvarc és földpátpótlók 0 %-os vonala mentén fordítanak egybe.

Ha egy háromszögdiagramban egy adott minta több időpontban vizsgált értékeit ábrázoljuk, akkor az ábrázolásmód alkalmas **időbeli folyamatok** elemzésére is. Példa erre az egyes országok foglalkoztatottsági szerkezetének vagy a szektorok által termelt GDP-nek időbeli módosulása.

A háromszögdiagramban történő adatábrázolás kissé eltér a szokásos két-dimenziós ábrázolásmódoktól, ugyanis a két független adattengely nem merőleges, nem a zérushelyen metszi egymás, és nem is látszódik az ábrán (azaz nem Descartes-féle koordináta-rendszert használunk). Egy-egy tengely a háromszög

egyik oldalfelezőpontjától a szemben levő csúcsig tart, a beosztása a következő: 0 % a háromszög oldalánál, 100 % a háromszög csúcsánál, az osztás egyenközű. Így egy adott százalékos érték az erre a tengelyre merőleges, a tengelyt a megfelelő értéknél metsző egyenesen található. Ha a három adatból kettőnek megkeresük az ilyen módon definiált egyenesét, akkor a két egyenes metszéspontja megadja az adatpont helyét a háromszögdiagramban.

A könnyebb ábrázolás és kiolvasás végett gyakran a legfontosabb százalékos értékek vonalait be is szokták húzni a diagramba (ezek általában 25 %-onként vagy 10 %-onként behúzott szakaszok). Mivel ezek a szakaszok elérik a diagram szélét, így a hozzájuk tartozó értékek a háromszög szélén feltüntethetők. Mivel egy-egy ilyen pontba két szakasz is befut, melyek eltérő paraméterhez tartoznak és eltérő százalékos értéket jelentenek, így ezek a felírások néha zavaróak is lehetnek. Ennek elkerülése végett általában azt a módszert szokták alkalmazni, hogy egy-egy oldalon csak egy paraméter szakaszainak értékét írják ki, általában egységesen a pozitív vagy negatív irányba haladva. Néha azonban ennek kivitelezése akadályokba ütközhet. E beosztásnak köszönhetően egy másik ábrázolási gyakorlat szerint a háromszög oldalain jelölt skála és az oldalakkal párhuzamos segédvonalak metszéspontja alapján is ábrázolható a minta.

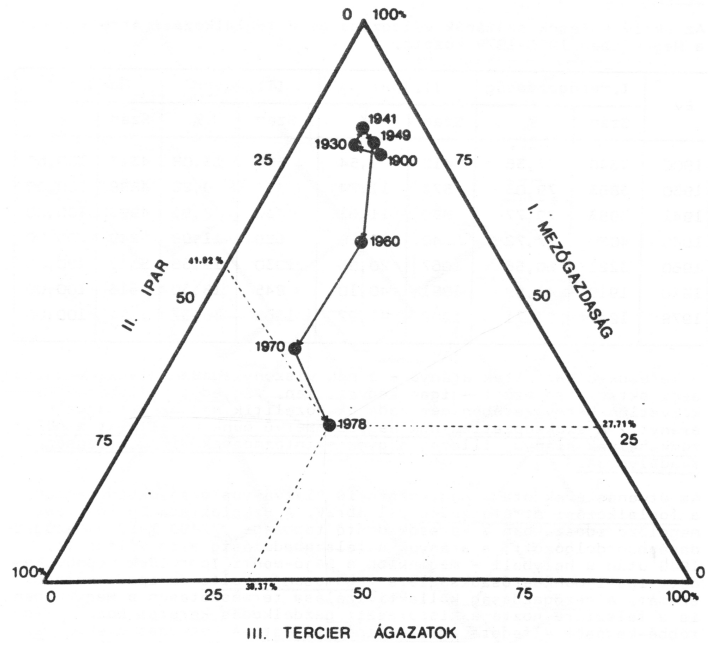
Több változó esetén (amennyiben összegük továbbra is állandó) többféleképpen járhatunk el. Az egyik, hogy térbeli tetraéder-diagramot használunk az elemzésekhez, de ennek megjelenítési lehetőségei korlátozottak. Másik lehetőség, hogy több változót összevonjunk (tercier és kvaternel aránya,  $\text{Na}^+$  és  $\text{K}^+$ -ionok összevonása a vízminták összetételének egyenértékszázalékos vizsgálatánál), esetleg adatredukcióval, változók összevonásával (lásd faktoranalízis).

A háromszögdiagramban a fenti összevonásokon túl egyéb összetett vagy származtatott értékek – így változók szorzatai, hányadosai – is előfordulhatnak. Ebben az esetben azonban fokozottabban kell figyelni, hogy a százalékképzés elvégezhető legyen (ennek egyik feltétele az azonos mértékegység). Azokban az esetekben, ahol a három adat jelentősen eltérő nagyságrendű, vagy a csoportképzés szükségessé teszi, a tényleges adatok helyett azok többszörösét vagy hányadosát hasonlítjuk össze a többi adattal.

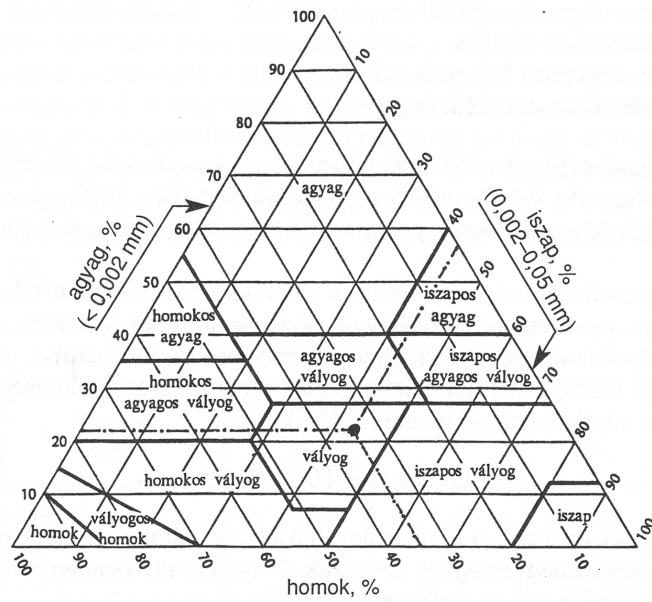
A háromszögdiagram Excelben is megszerkeszthető, jóllehet az alapfunkciók között nem szerepel. Mivel azonban a háromszögdiagram pontjai is értelmezhetők  $x; y$  koordinátaként, így a két rendszer között van átszámítási lehetőség:

$$x = a + \frac{b}{2} \quad \text{és} \quad y = \frac{b}{2} \cdot \sqrt{3},$$

ahol  $x$  és  $y$  a Descartes-féle koordináta-értékek, a és  $b$  pedig a háromszögdiagram két változója.

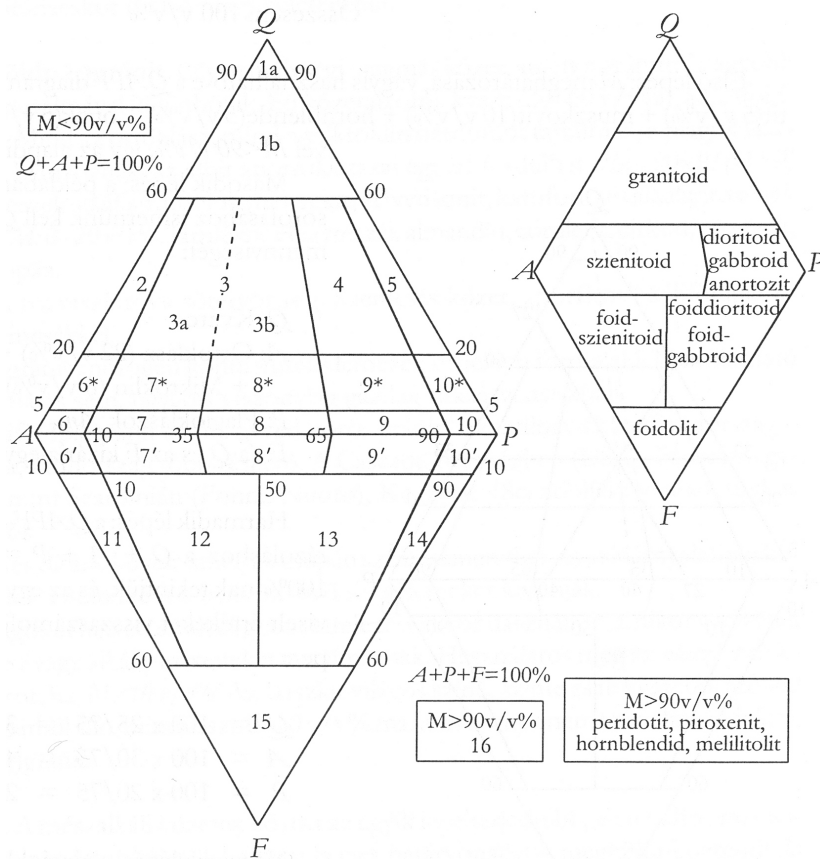


Példa időbeli változások (trendek) bemutatására háromszögdiagramban



Példa csoportalkotásra háromszögdiagramban  
(22% agyagfrakció, 35% homok, 43% iszap)





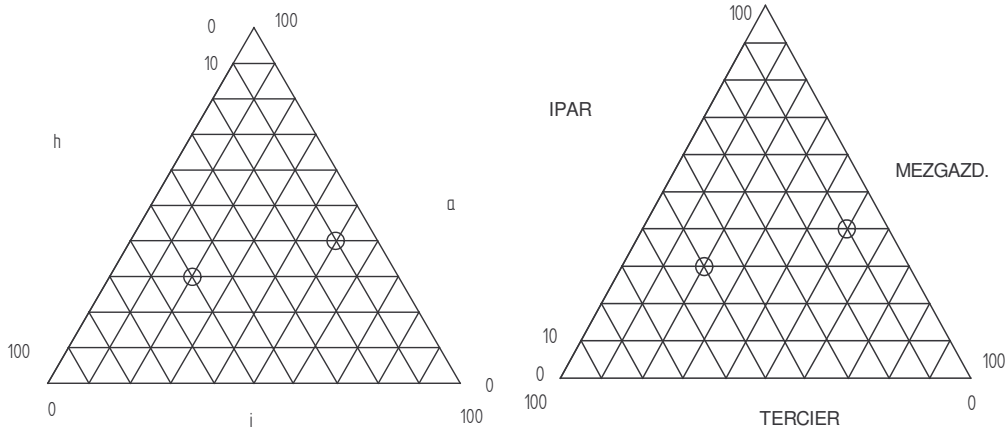
Streckeisen-diagram

### Feladatok

- 1, Határozza meg a két, körrel jelölt ország foglalkoztatási szerkezetét és helyezze el a mezőgazdaság=10%, ipar=40%, terciér=50% értékkel jellemezhető C országot a diagramon!
- 2, A fejlett országokra jellemző, hogy a mezőgazdaságban foglalkoztatottak aránya 10% alatti, a szolgáltató szektorban dolgozóké 50% feletti, míg a fejlődő országokra éppen a fordítottja jellemző. A volt szocialista országokra, illetve az ipari forradalom fázisában járó országokra jellemző, hogy az iparban dolgozók aránya nagyobb, mint 50%. Az említett adatok alapján rajzolja be a határvonalakat és alkoss csoportokat!
- 3, Piripócsón a fiatalok (0–18) aránya 1970-ben 40%-volt, a felnőtt korúak (18–65) aránya 35%, a nyugdíjasoké (>65) értelemszerűen 25%. 1980-ban ez rendre 30%, 40% és 30% volt, majd 1990-re az értékek rendre 20%-ra, 40%-ra és 40%-ra változtak. 200-ben viszont a fiatalok aránya 29%-ra, a fel-

nöttek 38%-ra a nyugdíjasoké 33%-ra nőtt. Ábrázolja háromszögdiagramon a változás irányát és adjon lehetséges magyarázatot a változás mögött álló társadalmi-gazdasági folyamatokra!

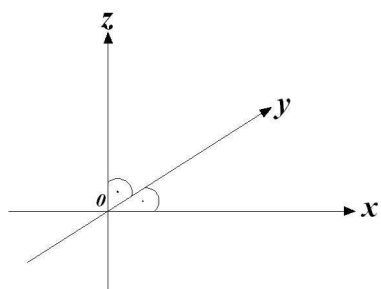
- 4, Egy talajminta 35%-os homok, 24%-os agyag és 41%-os iszaptartalommal rendelkezik. Minősítse a mintát az USDA rendszere alapján.



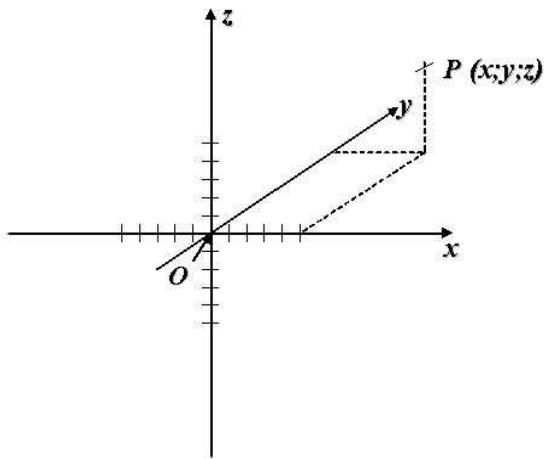
- 5, Határozza meg a háromszög-diagramon feltüntetett minták szemcseeloszlását, majd helyezd el a diagramon agyag=20%, iszap=30%, homok=50% mintát!
- 6, Egy talajminta 29% iszap, 33% agyag és 25%-os homoktartalommal bír. Ábrázolható-e ez a pont háromszögdiagramon?
- 7, QAPF- (Streckeisen-) diagramon ábrázolja és minősítse a következő értékekkel rendelkező mintát:  
 kvarc (Q) 25%, ortoklász és mikroklin (A), 8%, plagioklászok (P) 20%, f=0% (F és Q kizárják egymást!) Ekkor a  $Q+A+P=75$  v%, amit tekintünk 100%-nak és arányosan számoljuk vissza a %-értékeket Q-ra, A-ra, P-re.

### 1. 3. Háromdimenziós koordináta rendszerek

#### A Descartes-féle koordináta-rendszer



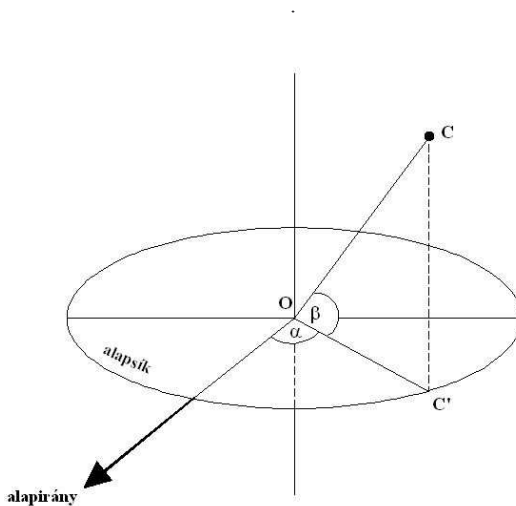
Vegyünk fel a síkon egy derékszögű Descartes-féle koordináta rendszert, majd a rendszer kezdőpontjában állítsunk merőleges számegeyenest a síkra. Így lehetőség adódik arra, hogy a térben kiválasztott pont helyét a három tengelytől mért merőleges távolsággal megadjuk.



A térbeli  $P(x; y; z)$  pont koordinátáinak meghatározásánál először a pontnak a síkra eső merőleges vetülete az előzőekben már megismert módon meghatározzuk a pont  $x$  és  $y$  koordinátáit a síkon. A harmadik koordinátát ( $z$ ) a  $z$  tengelyen mérjük a síktól mért távolsággal. Ennek megfelelően a pontot három számértékkel adhatjuk meg. A tér pontjai és a rendezett valós számhármas között kölcsönös és egyértelmű kapcsolat van.

A pont jelölése:  $P(x, y, z)$

### Polár térkoordináta-rendszer

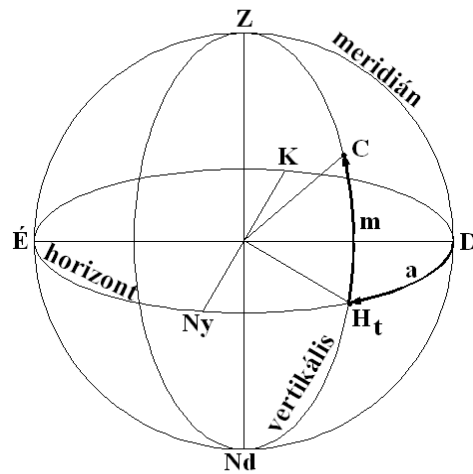


Vegyünk fel a síkon az álláspontunkban egy alapsíkra merőleges tengelyt és válasszunk ki a síkon egy alapirányt. A térbeli pont helyzetét egyértelműen két szög- és egy távolság-értékkel adhatjuk meg. A  $C$  pont koordinátáinak meghatározásánál először lemérjük az alapsíkon az alapiránytól a  $C$  pont merőleges vetületéig a szöget ( $\alpha$ ), majd az  $OC$  és a merőleges vetülete által bezárt szöget ( $\beta$ ) és végül meghatározzuk az  $OC$  távolságot. A pont jelölése:  $C(\alpha, \beta, d)$

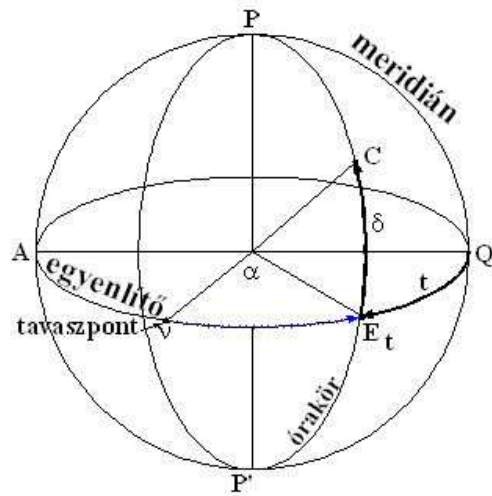
### Geográfiai alkalmazások

A térbeli koordináta-rendszereket elsősorban az égbolti tájékozódásnál használjuk. Ezeknek a rendszereknek az alapja a polár koordináta rendszer. Arra azonban ügyelnünk kell, hogy a csillagászatban az alapirány a **dél** és a szög mérését a különböző koordináta-rendszerekben eltérő módon végezzük:

- A **horizont égi koordináta rendszerben** az alapsík (a horizonton) az azimut ( $a$ ) szöget a déli iránytól az óramutató járásával megegyezően a horizontális talppontig, majd a magassági szöget ( $m$ ) a horizontális talpponttól a zenit irányába a csillagig mérjük. Az éggömbön a csillag helyének a kijelöléséhez ez a két koordináta elegendő, ugyanis a harmadik koordinátát (a csillag távolságát) itt nem vesszük figyelembe (a Földtől eltérő távolságra lévő csillagokat egy éggömbre vetítjük).

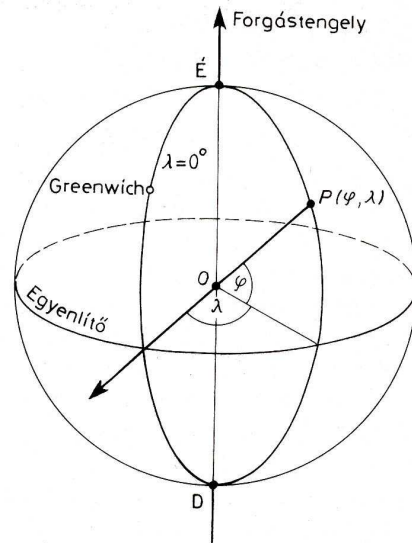


- Az **égi egyenlítői rendszerben** az alapsík az égi egyenlítő, a csillag helyét meghatározó egyik koordináta az óraszög ( $t$ ), a másik pedig a csillag deklinációja ( $\delta$ ). A harmadik koordinátával itt sem számolunk. Az óraszöget az égi egyenlítő és a meridián metszéspontjától ( $Q$ ) a csillag egyenlítői talppontjáig mérjük az óramutató járásával megegyezően. A deklinációt pedig az egyenlítői talpponttól a csillagig. A deklináció értéke lehet pozitív, vagy negatív attól függően, hogy a csillag az északi, vagy a déli félgömbön található. Az égi egyenlítői rendszerben arra is lehetőség van, hogy az óraszög helyett a csillag rektaszczenzióját ( $\alpha$ ) adjuk meg koordinátaként, akkor arra kell ügyelni, hogy a rektaszczenziót a tavaszponttól ( $v$ ) a csillag egyenlítői talppontjáig az óramutató járásával ellentétesen kell felmérnünk.

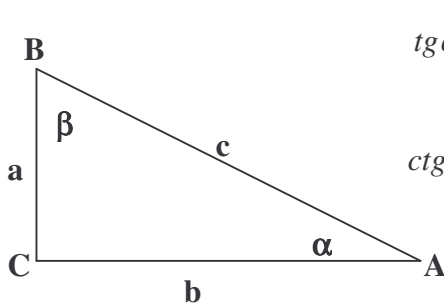


- Ha a polár térkoordináta-rendszert a **földgömbön** egy pont meghatározására akarjuk használni, akkor alapsíkul az egyenlítőt, alapiránynak pedig a meg-

állapodásnak megfelelően a Greenwichen áthaladó meridián ( $0^\circ$ ) és az egyenlítő metszéspontját választjuk. Az egyik koordináta a Föld középpontjából az adott felszíni ponthoz húzott sugár és az egyenlítő síkja által bezárt szög (szélességi szög =  $\phi$ ), a másik koordinátát pedig a kezdő meridiántól az adott pont egyenlítői vetületéig mérjük (földrajzi hosszúság =  $\lambda$ ). A földrajzi szélesség értéke számítási feladatoknál lehet pozitív, vagy negatív attól függően, hogy a felszíni pont az egyenlítőtől északra, vagy délre található. Egyébként a pontok koordinátáinak jelölésénél az égtáj (É, D) megjelölését használjuk. A hosszúsági értékeket a kezdő meridiántól keletre és nyugatra  $180 - 180^\circ$ -ig mérjük, és számítási feladatoknál itt is pozitív – negatív előjelet használunk, a koordináták jelölése az égtájak (K, Ny) megadásával történik.



#### 1. 4. Trigonometria



$$\operatorname{tg} \alpha = \frac{a}{b} \quad \operatorname{tg} \beta = \frac{b}{a} \quad \sin \alpha = \frac{a}{c} \quad \cos \alpha = \frac{b}{c}$$

$$\operatorname{ctg} \alpha = \frac{b}{a} \quad \operatorname{ctg} \beta = \frac{a}{b}$$

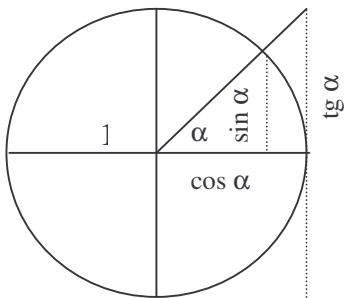
$\alpha$  és  $\beta$  pótszögek ( $\alpha + \beta = 90^\circ$ )

Egy szög szögfüggvényének értéke egyenlő pótszögének pótfüggvény-értékével!  
 {  $\sin(x)$ -nek  $\cos(x)$ ,  $\cos(x)$ -nek  $\sin(x)$ ,  $\operatorname{tg}(x)$ -nek  $\operatorname{ctg}(x)$  és  $\operatorname{ctg}(x)$ -nek  $\operatorname{tg}(x)$  }

$$\cos(90^\circ - \alpha) = \sin \alpha \quad \sin(90^\circ - \alpha) = \cos \alpha$$

$$\operatorname{tg}(90^\circ - \alpha) = \operatorname{ctg} \alpha \quad \operatorname{ctg}(90^\circ - \alpha) = \operatorname{tg} \alpha$$

$$\sin(180^\circ - \alpha) = \sin \alpha \quad \cos(180^\circ - \alpha) = -\cos \alpha$$



$$\operatorname{tg} \alpha = \frac{\sin \alpha}{\cos \alpha}$$

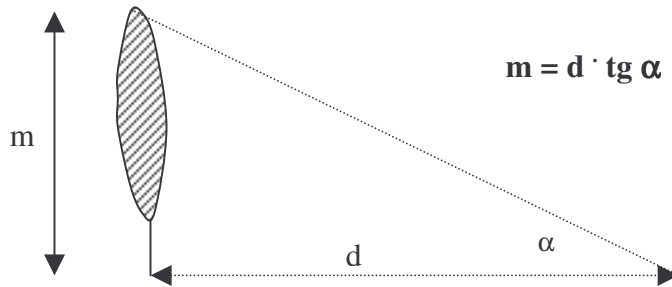
$$\sin^2 \alpha + \cos^2 \alpha = 1$$

Most már a szögfüggvények ismeretében a síkbeli polárkoordináták átszámíthatók Descartes derékszögű koordinátákra:

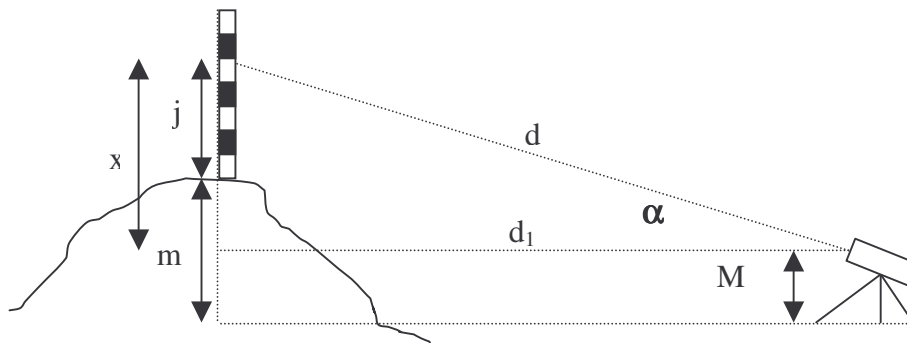
$$x = d \cdot \cos \alpha, \text{ illetve } y = d \cdot \sin \alpha$$

### Geográfiai alkalmazások

- Sík terepen a mérhető (megközelíthető) távolságban elhelyezkedő tereptárgyak (fa, épület, feltárás, stb.) magassága a távolság és a magassági szög ismeretében meghatározható.



- Domb- és hegyvidéki területen a terepi felméréseknél használt képleteket használjuk.



$$d_1 = d \cdot \cos \alpha$$

$d = 100 \cdot \mathbf{l}$  a  $d$  távolság kiszámításánál 100 a műszer szorzószáma,  $\mathbf{l}$  pedig merőleges rálátás esetén a műszer szátkeresztjének alsó – felső jelénél a lécről leolvasható értékek különbsége. Mivel a rálátás ferdeszögű, ezért a lécről leolvasott értékek különbségét  $\cos \alpha$ -val szorozni kell:  $\mathbf{l} = \mathbf{l}_1 \cdot \cos \alpha$

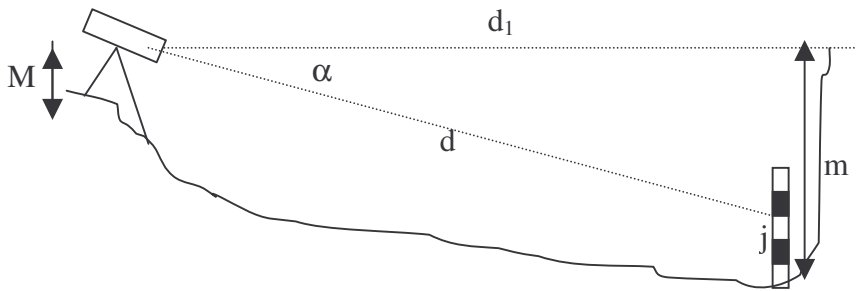
Ennek megfelelően  $d = 100 \cdot \mathbf{l}_1 \cdot \cos \alpha$

$$d_1 = 100 \cdot \mathbf{l}_1 \cdot \cos \alpha \cdot \cos \alpha \longrightarrow \mathbf{d}_1 = 100 \cdot \mathbf{l}_1 \cdot \cos^2 \alpha$$

$$\begin{aligned} \operatorname{tg} \alpha &= \frac{x}{d_1} \longrightarrow x = d_1 \cdot \operatorname{tg} \alpha; \quad m = x - j + M \\ m &= d_1 \cdot \operatorname{tg} \alpha - j + M; \quad m = 100 \cdot l_1 \cdot \cos^2 \alpha \operatorname{tg} \alpha - j + M \end{aligned}$$

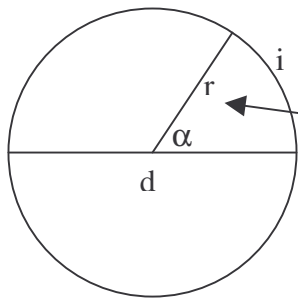
depressziós szög esetén:

$$m = 100 \cdot l_1 \cdot \cos^2 \alpha \operatorname{tg} \alpha + j - M$$



### 1. 5. Nevezetes vonalak

**Kör**



$$K = 2r\pi$$

$$T = r^2\pi$$

**Körcikk**

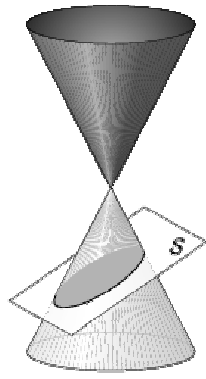
$$K = 2r \cdot i$$

$$i = \frac{2r\pi}{360^\circ} \cdot \alpha$$

$$T = \frac{r^2\pi}{360^\circ} \cdot \alpha^\circ \quad \text{vagy} \quad T = \frac{i \cdot r}{2}$$

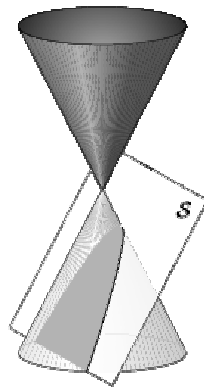
$$\text{Körszelet - ívhossz: } I = \frac{r \cdot \pi \cdot \alpha}{180^\circ} \quad \text{területe: } T = \frac{r^2}{2} \cdot \left( \frac{\alpha^\circ \cdot \pi}{180^\circ} - \sin \alpha^\circ \right)$$

**Kúpszeletek**



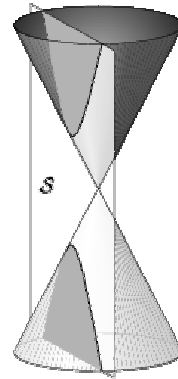
(a)

a, Ellipszis



(b)

b, Parabola



(c)

c, Hiperbola

**Egyenletük**

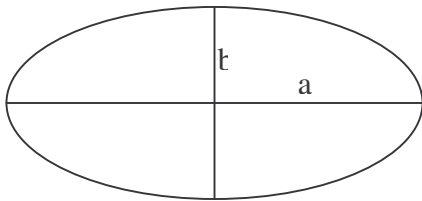
**ellipszis:**  $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$

**parabola:**  $y = a \cdot x^2 + b$

**hiperbola:**  $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$

**kör:**  $x^2 + y^2 = 1$

**egyenes:**  $y = m \cdot x + b$



**Ellipszis kerülete:**

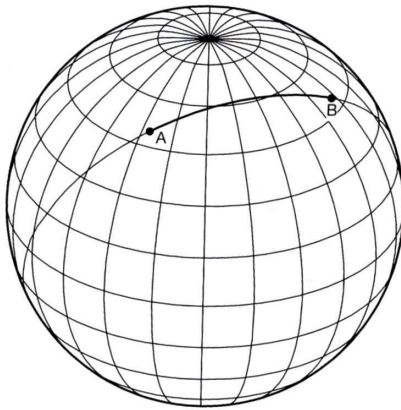
$$K = \pi \cdot \left( \frac{a+b}{2} + \sqrt{\frac{a^2+b^2}{2}} \right)$$

**Területe:**

$$T = a \cdot b \cdot \pi$$

ahol az „a” és „b” az ellipszis féltengelyei.

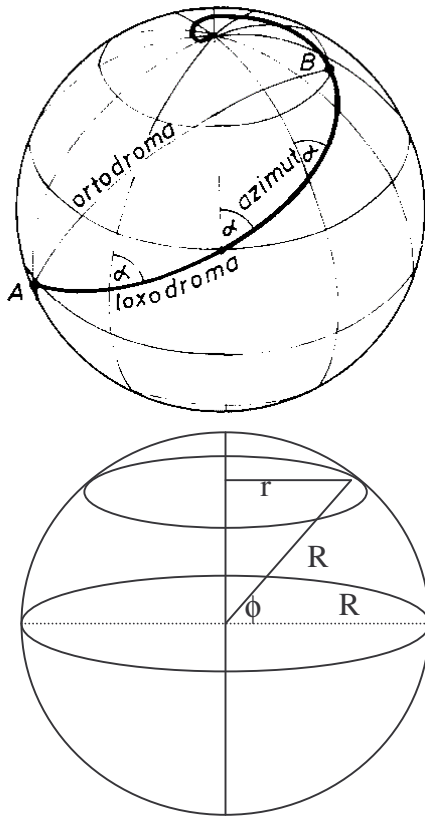
**Ortodróma**



A legnagyobb gömbi főkörökort ortodrómának nevezzük. Középpontjuk a gömb középpontja. Az ortodróma két pontja a kört két részre (ívdarabra) bontja. Ezek közül a rövidebb a gömbön a két pont között a legrövidebb távolságot adja.



## Loxodróma



Az a ferde futású gömbfelületi görbe vonal, amely a tengely végpontjaiban áthaladó ortodromákat mindig ugyanakkora szögek ( $\alpha$ ) alatt metszi.

### Geográfiai alkalmazások

A szélességi kör hosszának meghatározása

$$K = 2r\pi$$

$$r = R \cdot \cos \phi,$$

ahol  $r$  a szélességi kör  $R$  pedig a Föld sugara

$$K = 2R\pi \cos \phi$$

### Feladatok

- 1, Ábrázoljuk térbeli polár koordináta rendszerben azt a „nyugvó” csillagot, amelynek koordinátái: azimut =  $50^\circ$ , magasság =  $30^\circ$ !
- 2, Határozzuk meg annak az egyedülálló fának a magasságát, amelyet 20 méter távolságból  $30^\circ$  –os magassági szög alatt látunk!
- 3, Milyen magas az a kunhalom, amely mérésénél a következő adatokat kaptuk:  $M = 1,5$  m;  $l_1 = 0,4$  m;  $j = 1,80$  m;  $\alpha = 30^\circ$  ?
- 4, Egy település művelésági megoszlását 5 cm sugarú kördiagramon ábrázoljuk. Mekkora lesz a körcikkek szöge és területe, ha a szántó 70%, a kert+gyümölcsös 15%, a rét+legelő 5%, és az egyéb területek 10%-ban részesülnek?
- 5, Mekkora annak az ellipszis alakú parknak a területe, amelynek a hossza 80m, szélessége 25m?
- 6, Határozza meg Debrecen szélességén egy földrajzi fok hosszúságát!

- 7, Milyen távol van a Földön az a két hely, amelynek koordinátái:  
 $P_1(\text{É.sz.}32^\circ; \text{K.h. } 12^\circ 40')$  az egyik, és  $P_2(\text{É.sz.}32^\circ; \text{K.h. } 16^\circ 10')$ ;  
 $P_1(\text{É.sz.}32^\circ; \text{K.h. } 12^\circ 40')$  az egyik, és  $P_2(\text{É.sz.}32^\circ; \text{Ny.h. } 17^\circ 20')$ ?)
- 8, Milyen távol van a Földön az a két hely, amelynek koordinátái:  
 $P_1(\text{É.sz.}32^\circ 40'; \text{K.h. } 18^\circ)$  az egyik, és  $P_2(\text{É.sz.}32^\circ 10'; \text{K.h. } 18^\circ)$ ;  
 $P_1(\text{É.sz.}32^\circ 40'; \text{K.h. } 18^\circ)$  az egyik, és  $P_2(\text{D.sz. } 27^\circ 20'; \text{K.h. } 18^\circ)$ ?)
- 9, A 60. szélességi kör hossza hány %-a az Egyenlítőnek?
- 10, Határozzuk meg 1 földrajzi fok hosszúságát egy adott szélességi körön!

$$1 \text{ fok} = \frac{2R\pi \cdot \cos \phi}{360^\circ}$$

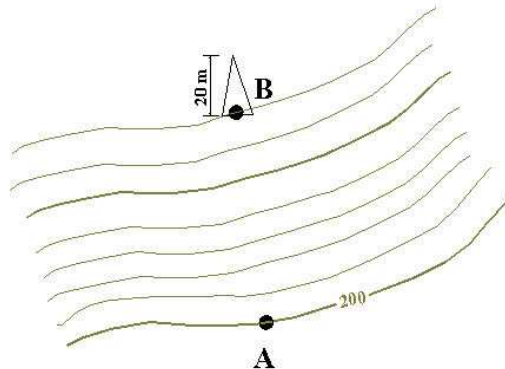
- 11, Határozzuk meg két olyan hely távolságát a Földön, amelyek azonos szélességi körön helyezkednek el!  $P_1(\phi_1; \lambda_1)$  az egyik,  $P_2(\phi_1; \lambda_2)$  a másik hely koordinátái, akkor a távolság fokban:  $\lambda_2 \pm \lambda_1$  (ha az egyik pont a kezdőmeridiántól Ny-ra a másik pedig K-re van, akkor kell az összeadást alkalmazni!)

$$\text{távolság} = \frac{2R\pi \cdot \cos \phi}{360^\circ} \cdot (\lambda_2 \pm \lambda_1)$$

- 12, Határozzuk meg két olyan hely távolságát a Földön, amelyek azonos hosszúsági körön helyezkednek el!  $P_1(\phi_1; \lambda_1)$  az egyik,  $P_2(\phi_2; \lambda_1)$  a másik hely koordinátái, akkor a távolság fokban:  $\phi_2 \pm \phi_1$  (ha az egyik pont az egyenlítőtől É-ra a másik pedig D-re van, akkor kell az összeadást alkalmazni!)

$$1 \text{ fok} = \frac{2R\pi}{360^\circ}, \quad \text{távolság} = \frac{2R\pi}{360^\circ} \cdot (\phi_2 \pm \phi_1)$$

- 13, Mekkora a felderítő (A) által rajzolt térképvázlat méretaránya, ha a vázlaton látható 20 m magas torony csúcsa 10 fokos szög alatt látszik (szintvonalak 10 m-enként)? Mekkora a valós távolság A és B pontok között és mekkora a térszín átlagos lejtése?



## 2. Gömbháromszögtani tételek és alkalmazásuk. Távolság és terület mérésének lehetőségei, számítási feladatok.

### 2. 1. Gömbi geometria

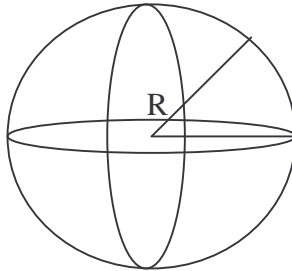
#### Gömb

A gömb olyan pontok mértani helye a térben, amelyek egy ponttól, a gömb középpontjától egyenlő távolságra vannak. Ha egy félkört a végpontjait összekötő szakasz (átmérő) körül forgatunk, akkor egy oly felületet kapunk, amelynek minden pontja az átmérő felező pontjától egyenlő távolságra van. A távolság a gömb sugara ( $R$ ).

$$\text{Felszíne: } F = 4R^2 \cdot \pi \quad \text{Térfogata: } V = \frac{4R^3 \pi}{3}$$

Egy sík a gömböt két gömbszeletre bontja.

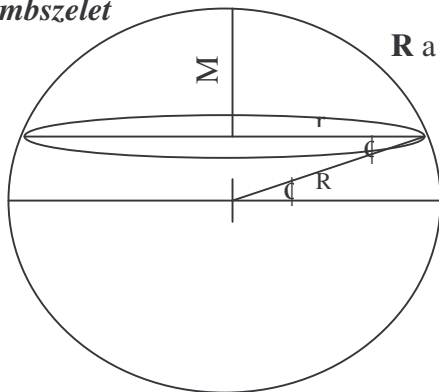
A gömb középpontján áthaladó sík és a gömbfelület metszetét főkörnek (ortodrómának) nevezzük. A főkörív a főkör egy szakasza. A főkörív középponti szöge a főkörív két végpontjából húzott gömbi sugarak által bezárt radiánban mért szög. ( $360^\circ = 2\pi$  radian)



A főkörív hossza (két pont távolsága) a pontokon áthaladó főkör rövidebb ívének hossza (Két pont között a legrövidebb út.). A főkörívek által bezárt szöget a síkjaik hajlásszögével mérünk.

A gömbkétszög két gömbi főkör által határolt terület. Két főkör a gömböt négy gömbkétszögre osztja. Felszíne  $F = 2R^2\alpha$  (a két főkör szöge radiánban egyértelműen meghatározza). (A teljes gömbfelszín  $4R^2\pi$ .)

#### Gömbszelet



$R$  a gömb,  $r$  a gömbszelet alapkörének a sugara

$$\text{Felszín: } A = T_{\text{kör}} + \text{Palást}$$

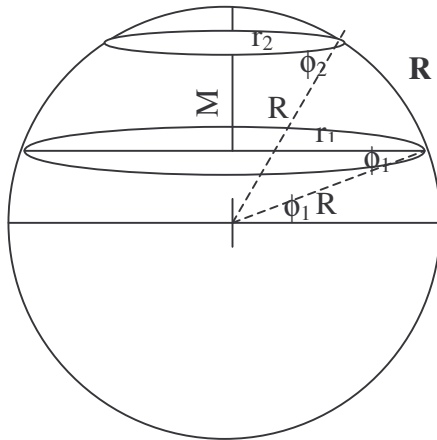
$$A = r^2\pi + 2R\pi \cdot M, \quad \text{ahol}$$

$$M = R - R \cdot \sin \phi = R \cdot (1 - \sin \phi)$$

$$A = r^2\pi + 2R^2\pi \cdot (1 - \sin \phi)$$

$$\text{Térfogat: } V = \pi \cdot M^2 \left( R - \frac{M}{3} \right) = \pi \cdot \frac{M}{6} (M^2 + 3r^2)$$

### Gömbön (gömbreteg)



$R$  a gömb,  $r_1$  és  $r_2$  a gömböv köreinek a sugara

**Felszín:**  $A = T_{1\text{kör}} + T_{2\text{kör}} + \text{Palást}$

$$T_{1\text{kör}} = r_1^2 \pi, \quad T_{2\text{kör}} = r_2^2 \pi$$

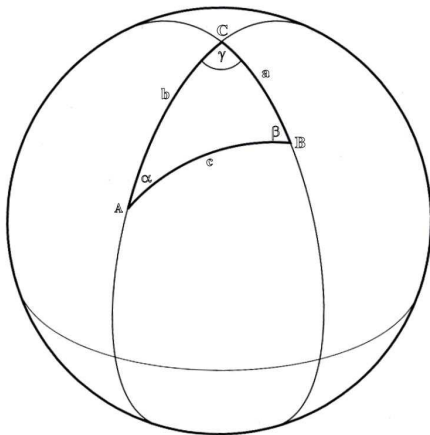
$$P = 2R\pi M$$

$$M = R \sin \phi_2 - R \sin \phi_1 = R (\sin \phi_2 - \sin \phi_1)$$

$$P = 2R^2 \pi (\sin \phi_2 - \sin \phi_1)$$

**Térfogat:**

$$V = \frac{\pi \cdot M}{6} (M^2 + 3r_1^2 + 3r_2^2)$$



A gömbháromszög a gömbfelület három pontját összekötő,  $\pi$ -nél rövidebb három fő-körív által határolt terület. Szögei kisebbek  $\pi$ -nél, szögeinek összege viszont nagyobb  $\pi$ -nél. Felszíne:  $F = R^2(\alpha + \beta + \gamma - \pi)$ .

## 2. 2. Gömbháromszögtani tételek

A gömbháromszög oldalait és szögeit a cosinus- és sinus tételekkel határozhatjuk meg.

### **Cosinus-tétel:**

Egy gömbháromszög oldalának cosinusát megkapjuk, ha a másik két oldal cosinusának szorzatához hozzáadjuk e két oldal sinusának és a közbezárt szög cosinusának szorzatát.

Képlet formában:

$$\cos a = \cos b \cdot \cos c + \sin b \cdot \sin c \cdot \cos \alpha$$

$$\cos b = \cos a \cdot \cos c + \sin a \cdot \sin c \cdot \cos \beta$$

$$\cos c = \cos a \cdot \cos b + \sin a \cdot \sin b \cdot \cos \gamma$$

### **Sinus-tétel:**

Egy gömbháromszögben két oldal sinusának aránya egyenlő az oldalakkal szemben fekvő szögek sinusának arányával.

Képlet formában:

$$\frac{\sin a}{\sin b} = \frac{\sin \alpha}{\sin \beta} \quad \frac{\sin b}{\sin c} = \frac{\sin \beta}{\sin \gamma} \quad \frac{\sin a}{\sin c} = \frac{\sin \alpha}{\sin \gamma}$$

### **Geográfiai alkalmazások**

A gömbháromszögekre levezetett cosinus- és sinus-tétel jól alkalmazható az éggömbi (csillagászati) feladatok megoldásánál. Abban az esetben, ha műszerekkel meghatározzuk egy égi objektum (pl. csillag) horizont koordináta rendszerben mérhető adatait, akkor az alábbi képletekkel az objektum egyenlítői rendszerbeli koordinátái kiszámíthatók.

#### **A csillagászati háromszög cosinus tételének képlete:**

$$\cos(90^\circ - \delta) = \cos(90^\circ - m) \cdot \cos(90^\circ - \phi) + \sin(90^\circ - m) \cdot \sin(90^\circ - \phi) \cdot \cos(180^\circ - a)$$

$$\sin(\delta) = \sin(m) \cdot \sin(\phi) - \cos(m) \cdot \cos(\phi) \cdot \cos(a), \text{ ahol}$$

( $\delta$ ): a csillag deklinációja, (m): a csillag magassága a horizont felett, (a): a csillag azimut értéke, amelyet a délponttól mérünk a csillag horizontális talppontjáig és

( $\phi$ ): az adott hely földrajzi szélessége.  $m = 0$  feltételt alkalmazva kiszámítható a kelő és nyugvó égitest azimutja (a déli kezdőirányhoz képest):

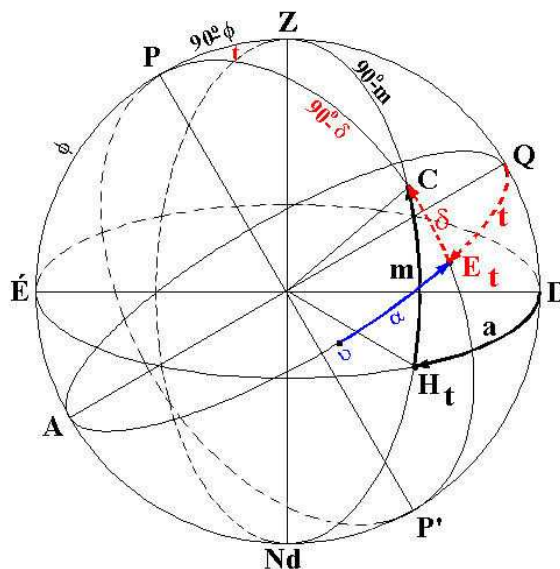
$$\cos(a) = -\sin(\delta) / \cos(\phi)$$

#### **A csillagászati háromszög sinus tételének képlete:**

$$\sin(90^\circ - \delta) : \sin(90^\circ - m) = \sin(180^\circ - a) : \sin t$$

$$\cos \delta : \cos m = \sin a : \sin t$$

$$\sin t = \frac{\sin a \cdot \cos m}{\cos \delta}, \text{ ahol } t \text{ a csillagidő.}$$



A tételek segítségével az egyenlítői rendszer adataiból a csillag horizont rendszerbeli adatai is meghatározhatók. Ekkor a képlet az alábbiak szerint módosul:

$$\sin m = \sin(\delta) \cdot \sin(\phi) + \cos(\delta) \cdot \cos(\phi) \cdot \cos t, \text{ illetve}$$

$$\sin a = \frac{\cos \delta \cdot \sin t}{\cos m}$$

A Nap magasságának és a deklinációjának ( $\delta$ ) ismeretében a földrajzi szélesség meghatározható:

$$\phi_1 = 90^\circ - m + \delta, \text{ ha a Nap a zenittől délre delel}$$

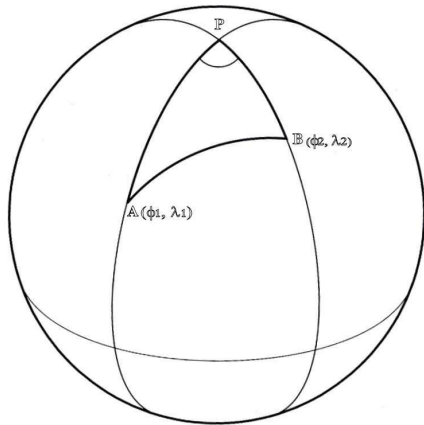
$$\phi_1 = -90^\circ + m + \delta, \text{ ha a Nap a zenittől északra delel.}$$

A nappal hosszát is a fentiek figyelembe vételével határozhatjuk meg. Napkeltétől ( $m = 0$ ) delelésig legyen az óraszög jele 't'. Ekkor

$$\cos t = -\operatorname{tg}\phi \cdot \operatorname{tg}\delta \quad (360^\circ = 24 \text{ h}, 15^\circ = 1 \text{ óra}, 1^\circ = 4 \text{ perc})$$

A nappal teljes hossza:  $2t$ . (Mivel  $t$  óraszög nem más mint a napkeltétől a delelésig eltelt idő, a nappal hossza  $2t$ , így a napkelte ideje:  $12:00-t$ ; a napnyugta  $12:00+t$  helyi idő szerint)

Abban az esetben, ha a Földet gömbnek tekintjük, akkor a felszíni pontok közötti távolságok könnyen kiszámíthatók.



$$P(0^\circ, 90^\circ)$$

$$A(\varphi_1, \lambda_1) \quad B = (\varphi_2, \lambda_2)$$

$$\cos c = \cos a \cdot \cos b + \sin a \cdot \sin b \cdot \cos \gamma$$

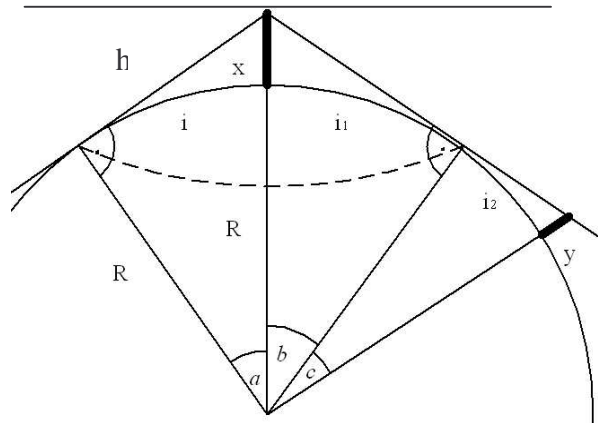
$$\cos AB = \cos(90^\circ - \varphi_2) \cdot \cos(90^\circ - \varphi_1) + \sin(90^\circ - \varphi_2) \cdot \sin(90^\circ - \varphi_1) \cdot \cos(\lambda_2 - \lambda_1)$$

ebből

$$\cos AB = \sin \varphi_2 \cdot \sin \varphi_1 + \cos \varphi_2 \cdot \cos \varphi_1 \cdot \cos(\lambda_2 - \lambda_1)$$

### Példák:

- 1, Milyen messzire lát el egy 2 m magas ember? Látható-e egy 50 m magas toronyból a 20 km-re levő 10 m magas torony?



Megoldás:

A látótávolság (a látóhatár, a horizont: szaggatott vonal, a hozzá tartozó „sugár” egy ívhossz:  $i$ ) meghatározása függ a szemlélő magasságától ( $x$ ). Az ábráról leolvasható, hogy  $i$  akkor határozható meg, ha ismert a hozzá tartozó középponti szög. Ekkor  $i/2R\pi = a/360$ . Az  $a$  meghatározható, mert  $R/(R+x) = \cos a$ , hiszen a horizont egy érintővel jelölhető ki, amely merőleges  $R$  fűldsugárra (6371 km).

Tehát egy 2 m magas ember  $i$  távolságra lát el, amihez  $a$  középponti szög tartozik:  $6371/6371,002 = \cos a$  (feltételezzük, hogy nincs terepakadály, nem számolunk a refrakció jelenségével és a Linke-féle homályossági tényezővel). A

tizedesjegyeket ne hanyagoljuk el, amikor kiszámoljuk a szöveget! Ez az  $a$  szög egyébként megegyezik a horizont-depresszió  $h$  szögével (merőleges szárú szögek).  $A = 0,0454^\circ$ , így  $i/2 = 6371 \cdot 3,14 = 0,0454/360$ , ahonnan  $i = 5,04$  km.

Előfordulhat, hogy álláspontunkról/magaslati kilátópontunkról egy  $y$  magasságú objektumnak csak a tornyát látjuk – ez azt jelenti, hogy az objektum messzebb van, mint a tényleges horizont (ahol az égbolt a földfelszín éri), de magassága miatt a nagy távolság ellenére is a horizont fölé ér. Ebben az esetben az  $i_1$  ívhossz megnövelendő  $i_2$  ívhosszal, ami nem más, mint az  $a$  távolság, amilyen messze az  $y$  magasságú objektumból ellátnánk, ha ott lennénk. Tehát  $b$  és  $c$  szög és a hozzá tartozó két ívhossz számítandó ki, a már említett módon.

Figyelembe kell azonban venni, hogy a két pont nem feltétlenül helyezkedik el ua. tengerszint feletti magasságon. Ha például álláspontunk tszf. 1000 méteren van, az  $y$  magasságú objektum viszont 200 méteren, akkor az  $R$  érték módosítandó, új vonatkoztatási felület jelölhető ki: ez az  $R+0,2=R_{új}$ . Az első esetben  $(R+0,2)+0,8$  álláspontunk magassága, míg a másik objektum teteje  $(R+0,2)+y$  magasságban van. Tehát  $\cos b = R_{új}/(R_{új}+0,8)$  és  $\cos c = R_{új}/(R_{új}+y)$ . Kis magasságok (magasságkülönbségek) esetén az áttérés az új vonatkoztatási felületre elhagyható.

Ha távolság az adott (mint most a 20 km), akkor azt a távolságot kell kiszámolni, ahonnan az adott  $y$  magasságú objektum még éppenhogy látszik. Ha ez nagyobb, mint 20 km, akkor az objektum 20 km-ről is látszik, ha 20 km-nél kevesebb a maximális láthatósági távolsága (pl. 16 km), akkor nem fog látszani 20 km-ről az objektum.

2, Repülőgépünk lezuhan és túlélőként ismeretlen helyen landolunk. Összes vagyunk egy karóra. Határozzuk meg álláspontunk földrajzi koordinátáit! Tétélezzük fel, hogy amikor ismeretlen álláspontunkon a Nap delel (azaz helyi idő szerint 12:00 van), a budapesti zónaidőt mutató karóránk 18:00-t mutat. Június 22-e van, a Nap delelés magassága  $60^\circ$  és északi irányban látszik!

Megoldás:

Mivel a Föld Ny–K irányban forog tengelye körül, így bizonyos, hogy Budapesttől nyugatra vagyunk, hiszen álláspontunkon a Nap éppen delel, amikor Budapesten már a horizont felé közelít. A Föld közel 24 h alatt fordul  $360^\circ$ -ot, egy óra alatt tehát  $15^\circ$  a szögelfordulása. Ez 6 óra alatt  $90^\circ$ - jelent. Mivel Bp. a keleti hosszúság megközelítőleg  $20^\circ$ -án helyezkedik el, álláspontunk a nyugati hosszúság  $70^\circ$  -án található.

A szélesség meghatározásához a dátum ismeretéből és a Nap delelés magasságából indulhatunk ki. A helyi dél megállapítható gnómon (földbe szúrt bot) segítségével (a legrövidebb árnyék jelöli a delet), ekkor a delelés magasság meghatározható. Ismeretes továbbá, hogy a dátum (ill. a Nap merőleges delelésé-



nek helye, deklináció,  $\delta$ ) és a Nap delelés magassága (m) alapján a szélesség meghatározható:

$$\varphi = 90 - m + \delta \text{ ill. } \varphi = -90 + m + \delta$$

Június 22-én a Nap a Ráktérítőn delel merőlegesen, tehát deklinációja  $23,5^\circ$ . A delelés magasság  $60^\circ$  (ezt Jákob botjával mérhetjük meg), így

$$\varphi = 90 - 60 + 23,5 \text{ ill. } \varphi = -90 + 60 + 23,5$$

$\varphi$  tehát  $53,5$  vagy  $-6,5^\circ$ . Példánkban a Napot északi irányban látjuk delelni amely éppen a Ráktérítő szélességén helyezkedik el így álláspontunk a Ráktérítőtől délre van, azaz a  $\varphi = -6,5^\circ$  a helyes megoldás. (Vagy kivárjuk az éjszakát és megkeressük a Sarkcsillagot: ha nem látjuk, akkor a déli féltekén vagyunk).

Ha a képletből a dátum és a szélesség ismert, akkor a Nap delelés magassága, ha pedig a szélesség és a delelés magasság ismert, akkor a dátum is meghatározható. Nemcsak „nevezetes” dátum alkalmával, hanem bármikor meghatározható a deklináció (éggömbi szélesség, a Nap merőleges delelésének helye), tudván, hogy a Nap látszólagos évi járása sinus-os.

A hosszúság is pontosítható, hiszen valójában egy helyi időt és egy zónaidőt vetettünk össze, ami nem teljesen helyénvaló. A budapesti időzóna központi hosszúsági köre a  $15^\circ$  (K.h.  $7,5^\circ - 22,5^\circ$ -ig tart a zóna), ennek a hosszúsági körnek a helyi idejét terjesztették ki az egész időzónára. Amikor itt tehát 18:00 van, akkor Budapesten a helyi idő már meghaladta a 18:00-t, hiszen a város megközelítőleg a 20. hosszúsági körön helyezkedik el ( $1^\circ = 4$  perc, azaz kb. 20 percről van szó). Valójában tehát a 18:00 órát és a 90 foknyi különbséget a 15 foktól kell mérni, így a hosszúság helyesen Ny.h.  $75^\circ$ .

### ***Feladatok***

- 1, Határozza meg a Föld felszínét és térfogatát, ha az egyenlítői sugara kerekítve: 6378 km!
- 2, Mekkora a Föld felszínének területe a 60. szélességi terület és a pólus között?
- 3, Mekkora a Föld felszínének területe a 30. és 45. szélességi körök között?
- 4, Melyik két szélességi kör osztja három egyenlő részre a Föld felszínét?
- 5, Mekkora a csillag deklinációja ( $\delta$ ) és óraszöge (t), ha Debrecenben ( $\varphi = 47,5^\circ$ ) a csillag azimutja (a)  $30^\circ$ , magassága (m)  $45^\circ$ ?
- 6, Debrecenben ( $\varphi = 47,5^\circ$ ) éppen zenitben (Z) delel egy csillag. Mennyi az óraszöge (t) és a deklinációja ( $\delta$ )?
- 7, Delelhet-e bármikor Magyarország valamely pontján magasabban a Nap, mint az Egyenlítőn?
- 8, Mikor és hol kel és nyugszik a Nap Debrecenben ( $\varphi = 47,5^\circ$ ) a nevezetes csillagászati napokon?

*(Figyelem! A kelés helyének, azimutjának kiszámításához  $m=0$  peremfeltételt alkalmazunk, hiszen a kelő és nyugvó égitest magassága a horizont felett ennyire, ha nem számolunk a refrakcióval! A kapott szögérték a déli irányból mérendő fel!)*

- 9, Milyen hosszú a nappal Debrecenben ( $\phi=47,5^\circ$ ) március 1.-én? Mikor és hol kel és nyugszik?
- 10, Milyen távol van Makó Jeruzsálemtől (a közhiedelemmel ellentétben a mondasban nem a település, hanem Makó vitéz szerepel...)?
- 11, Árpád apánk a honfoglalás után körbenézett, és hű vitézét, aki különösen nagy tehetséget árult el a szlávok lábasjóságainak és asszonyainak abajgatása terén, megjutalmazta. „Tied minden, ameddig a szem ellát” – imígyen szóla. Feltételezve, hogy hű harcosa 1,7 m magas volt és nem akadályozza semmi a kilátást (ugyanakkor elhanyagolva a refrakciót), mekkora területet kapott a harcos?  
*(Figyelem! A felszín görbült, tehát a gömbsüveg palástja a keresett terület!)*
- 12, Milyen messzire látni a Kékes-tetőről (refrakció és a kilátást zavaró tényezők elhanyagolva, a Linke-féle homályossági tényező minimális)?
- 13, Látszik-e a Csomolungmáról a tenger?
- 14, Látszik-e a Kékesről a szegedi dóm teteje (80 m) ?  
*(Figyelem! Szeged nem a tengerszinten helyezkedik el, tehát a Kékes-tető magassága a felszín felett nem 1014 m!)*
- 15, Számolja ki az É.sz.  $20^\circ$  és É.sz.  $30^\circ$  közötti területet és a térfogatot!
- 16, Hol hosszabb az éjszaka? Londonban (50. északi szélesség) december 22-én vagy Quitóban (Egyenlítő) március 21-én? Hol és mikor kel föl a Nap?
- 17, Repülőgép szerencsétlenség áldozata Peches Pistike, lezuhan, de túléli. Összes vagyona, ami megúszta a katasztrófát, a karórája. Mikor álláspontján de-lel a Nap, a bp-i időt mutató óráján délelőtt 10 óra van. A Nap delelés-magassága 45 fok, március 1-e van. Határozza meg az álláspont földrajzi szélességét és hosszúságát!
- 18, Számolja ki Lima és Budapest távolságát! *(Eltérő féltekén vannak, Lima szélessége negatív,  $-\phi$ ! és hosszúság tekintetében is eltérő féltekén vannak!)*

### 3. A differenciál- és integrálszámítás alapjai

Az elmúlt évtizedekben a középiskolai matematika tanterve sokat változott és ennek köszönhetően a differenciál- és integrálszámítás témakörét nem mindig, illetve nem mindenütt tanították. Úgy gondoljuk, hogy a földtudomány egyes területein adódnak olyan problémafelvetések, illetve feladatok, amelyek megoldása e két témakör ismeretében könnyebbé válik. Ebben a rövid fejezetben nem célunk a differenciálszámítás és integrálszámítás részletes tárgyalása, csak azokra a legalapvetőbb alapismeretekre térünk ki, amelyek egyrészt segítséget nyújthatnak a téma részletesebb megismerésénél, másrészt bizonyos földtudományhoz kapcsolódó számítások (pl. görbe vonallal határolt parcella területének meghatározása) megértésénél.

### 3.1. Differenciálszámítás

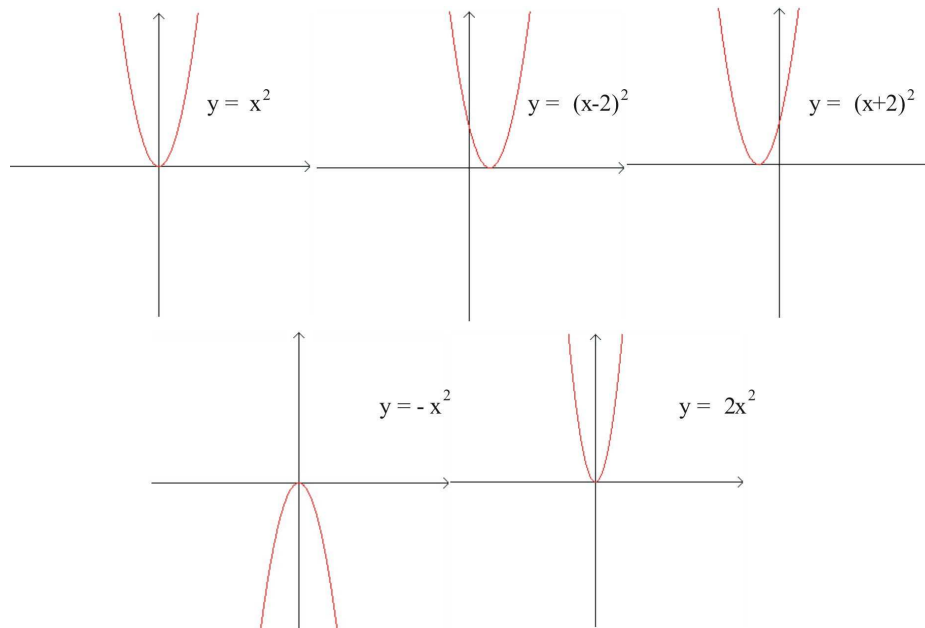
A téma egyszerűbb megértése érdekében először ismételjük át, a középiskolában már tanult, függvény fogalmát. A Descartes-féle derékszögű koordináta-rendszer tengelyeinek megszokott jelölését ( $x$ ,  $y$ ) használva azt mondjuk, hogy  $y$  az  $x$ -nek függvénye, ha  $x$  minden szóbjövő értékéhez  $y$ -nak legfeljebb egy meghatározott értéke tartozik. Azt, hogy  $y$  az  $x$ -nek függvénye általában  $y = f(x)$  jellel jelöljük. *(Természetesen a képlet betűi tetszőlegesen, de értelemszerűen változtathatók – pl.  $u = g(v)$ , ha  $a$   $v$  minden szóbjövő értékéhez  $u$ -nak meghatározott értéke tartozik.)*

Az  $x$ -et független változónak,  $y$ -t függő változónak szokás nevezni. A független változók ( $x$ ) alkotják a függvény értelmezési tartományát és az  $y$  megfelelő értékei pedig a függvény értékkészletét.

A függő változó értékeinek a függő változó értékeihez való rendelése sokféleképpen történhet. Gyakran egy képlet segítségével lehet  $x$  értékéből az  $f(x)$  értékét kiszámítani (pl. ha a függvény képlete  $y = x^2$  és a független változó  $x = 2$ , akkor a függvény értéke ezen a helyen  $y = 4$ ). Ha az  $x$  minden értékéhez ugyanaz az  $f(x)$  függvényérték tartozik, akkor a függvényt állandónak, konstansnak nevezzük (pl.  $y = f(x) = 5$  azt jelenti, hogy a független változó ( $x$ ) minden értékéhez 5-ös függvényérték tartozik).

A függvény geometriai szemléltetése (ábrázolása) derékszögű koordináta rendszerben történik, ahol a független változó értékei az  $x$  értékek és a függő változók, tehát a függvény értékei az  $y$  értékek (ha ilyen tengelyjelöléseket használunk). A függő változó összes értékének ábrázolása adja a függvény képét. A fentebb említett két függvény közül az  $y = f(x) = x^2$  képe parabola, az  $y = f(x) = 5$  képe pedig az  $x$  tengellyel párhuzamos egyenes, amely az  $x$  tengelytől 5 egységnyi távolságra van. A függvények képe a képlet változtatásával azonnal változik. Ha pl. a parabola képletét  $y = (x-2)^2$ ;  $y = (x+2)^2$ ;  $y = 2x^2$ ;  $y = -x^2$  alakúra változtatjuk, akkor a kép az alábbi ábrán látható módon változik. Ekkor azt mondjuk, hogy a függvényt transzformáltuk. A középiskolában a függvények különböző

típusait (elsőfokú, másodfokú, exponenciális, trigonometrikus, stb.) ismerhették meg.



Az  $y = f(x)$  függvény grafikus ábrázolásánál nem mindig kapunk összefüggő (folytonos) görbe vonalat. Ha egyszerűen akarjuk megfogalmazni a függvény folytonosságát, akkor azt mondjuk, hogy az  $x$  értékeinek kis változásánál a függvényérték  $y = f(x)$  is csak kevéssel változik. A folytonosság fogalmának pontosabb meghatározásához először tisztázni kell, hogy mit jelent az  $x$  és  $y$  értékeinek kis változása. Ennek érdekében röviden foglalkoznunk kell a függvény határértékével.

Az  $y = f(x)$  függvény **H** határértékét, vagy limesét (a továbbiakban rövidítve **lim**) az  $x_0$  pontban így jelölik:

$$H = \lim_{x \rightarrow x_0} f(x) \quad \{\text{kiolvasva } x \text{ tart } x_0\text{-hoz } f(x)\}$$

amelyen azt értik, hogy  $|H - f(x)|$  (vagyis az  $f(x)$  függvény eltérése a  $H$  határértéktől) kisebb, mint egy előre megadott tetszőlegesen kicsi pozitív  $\epsilon$  szám. Ez a feltétel ( $|H - f(x)| < \epsilon$ ) elérhető, ha  $x$  kevesebbel különbözik  $x_0$ -tól, mint egy megfelelően választott  $\delta$  szám, azaz, ha  $|x - x_0| < \delta$ . Tehát ez a  $\delta$  szám az előre megadott  $\epsilon$  értékétől függ, ezért gyakran  $\delta = \delta(\epsilon)$ -nal jelölik.

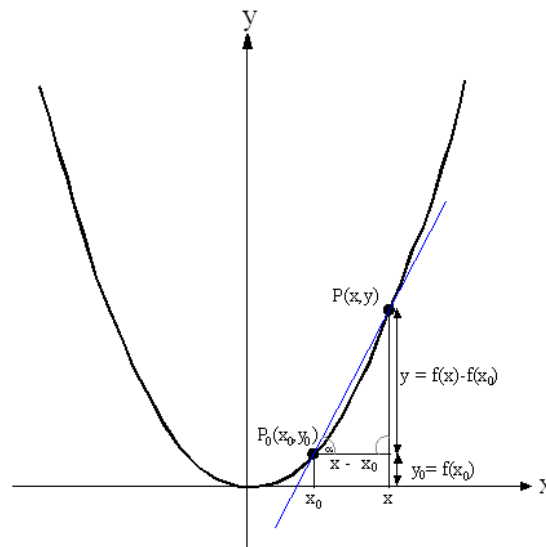
Ennek a meghatározásnak az a feltétele, hogy az  $f(x)$  függvény az  $x_0$  körüli intervallumban értelmezve legyen. Az  $\epsilon$  definíciójánál megkövetelt

$x \rightarrow x_0$  közelítésénél figyelembe kell venni az  $x - x_0$  előjelét, mert ettől függően lehet bal-, vagy jobboldali határértékről beszélni. Csak akkor beszélhetünk teljes határértékről, ha mindkét oldali határérték létezik, és azok egyenlők egymással ( $H_b=H_j$ ).

Most már a függvény folytonosságára vonatkozó definíció pontosan megfogalmazható, amely szerint az  $y = f(x)$  függvény akkor folytonos egy  $x_0$  pontban, ha az  $x$  tengely  $x_0$  pontja és ennek valamely környezete az értelmezési tartományhoz tartozik, továbbá az  $x_0$  pontban a függvénynek van bal- és jobb oldali határértéke, és azok megegyeznek a függvény  $x_0$  pontjához tartozó helyettesítési értékével, azaz  $\lim_{x \rightarrow x_0} f(x) = f(x_0)$ .

Ábrázoljuk az  $y = f(x) = x^2$  függvényt és válasszuk ki két pontját. Ezeket jelöljük  $P_0(x_0, y_0)$  és  $P(x, y)$ -al. Az értelmezési tartomány két értékéhez ( $x, x_0$ ) tartozó függvényérték  $y = f(x)$  és  $y_0 = f(x_0)$ . Ha összekötjük egyenes szakasszal a függvény két pontját, akkor olyan derékszögű háromszöget kapunk, amelynek egyik befogója a függvényértékek különbsége  $\{f(x) - f(x_0)\}$ , a másik befogója pedig az abszcissa értékek különbsége  $\{x - x_0\}$ . Ezek a különbségek (differenciák) a két pont helyzetéből adódnak, tehát attól függenek. Ha a két különbség hányadosát képezzük, akkor kapjuk meg az  $f(x)$  függvény  $x_0$  pontbeli differenciahányadosát, amely képlet formájában a következő:

$$\operatorname{tg} \alpha = d(f, x_0) = \frac{f(x) - f(x_0)}{x - x_0} \quad x \neq x_0$$



Az így képzett hányados tulajdonképpen a két befogó hányadosa megegyezik a derékszögű háromszög  $P_0$  pontjánál lévő hegyesszög ( $\alpha$ ) tangensének

értékével. Mivel a derékszögű háromszög átfogója a két ponton áthaladó szelő irányával azonos, ezért azt is mondhatjuk, hogy a differenciahányados geometriai jelentése a szelő iránytangense.

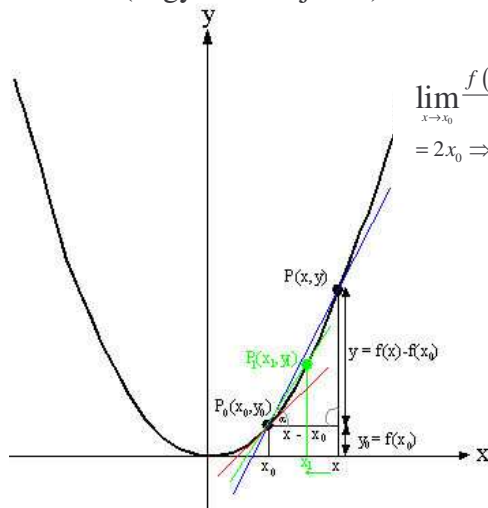
Ha az  $x \rightarrow x_0$  ( $x$  tart  $x_0$ -hoz), a független változók  $x_1, x_2, \dots, x_n$  értékéhez tartozó függvényértékek  $f(x_1), f(x_2), \dots, f(x_n)$  esetén az  $\{(x_0, f(x_0)), (x_1, f(x_1))\}, \{(x_0, f(x_0)), (x_2, f(x_2))\}, \dots, \{(x_0, f(x_0)), (x_n, f(x_n))\}$  pontpárok által meghatározott szelők iránytangensei

$\frac{f(x_1) - f(x_0)}{x_1 - x_0}, \frac{f(x_2) - f(x_0)}{x_2 - x_0}, \dots, \frac{f(x_n) - f(x_0)}{x_n - x_0}$  fokozatosan közelítik az

$\frac{f(x) - f(x_0)}{x - x_0}$  differenciahányados értéket, amely megadja a függvény

$(x_0, f(x_0))$  pontjába húzott érintőnek az iránytangensét. Az így kapott

$\text{tg } \alpha = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0}$  határértéket az  $f(x)$  függvény differenciálhányadosának (vagy deriváltjának) nevezzük.



A parabola esetén

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{x \rightarrow x_0} \frac{x^2 - x_0^2}{x - x_0} = \lim_{x \rightarrow x_0} \frac{(x - x_0)(x + x_0)}{x - x_0} = \lim_{x \rightarrow x_0} (x + x_0) = 2x_0 \Rightarrow (x^2)' = 2x$$

Ez a jelölés azt is jelenti, hogy a határérték attól az  $x_0$  helytől függ, amelyben számítjuk, és így  $x$ -nek a függvénye. Ha  $x_0$  bármely tetszőleges helyet jelölhet, akkor a függvénynek bárhol lehet határértéke. Ez az eset csak akkor áll fenn, ha az  $f(x)$  függvény  $x$ -ben folytonos. Szakadási helyen a függvény nem differenciálható. A differenciálhatóságból tehát következik a folytonosság, de ennek a fordítottja nem mindig igaz.

A fenti ábrán jól látszik, hogy ha az  $x$  tart  $x_0$ -hoz, akkor a görbén a  $P$  pont a  $P_0$  felé mozog, és a szelő iránya egyre jobban megközelíti az érintő irányát. Te-

hát az  $y = f(x)$  függvény differenciálhányadosának az  $y = f'(x)$ -nek a geometriai jelentése az hogy függvény  $(x, f(x))$  pontjába húzott érintő iránytangense  $f'(x)$ -xell egyenlő. A differenciálhányadost szokták még  $\frac{\partial f}{\partial x}$  vagy  $\frac{dy}{dx}$ -xel is jelölni.

Az egyszerű függvények differenciálhányadosát sok esetben a határértékekkel lehet kiszámítani. A feladatok megoldásában segít az alábbi táblázat.

$f(x)$	$f'(x)$	$f(x)$	$f'(x)$
c (konstans)	0	$e^x$	$e^x$
$x^n$	$n \cdot x^{n-1}$	$a^x$	$a^x \cdot \ln a$
$\sin x$	$\cos x$	$\ln x$	$\frac{1}{x}$
$\cos x$	$-\sin x$	$\lg x$	$\frac{1}{x} \cdot \lg e$
$\operatorname{tg} x$	$1 + \operatorname{tg}^2 x$	$\arcsin x$	$\frac{1}{\sqrt{1-x^2}}$
$\operatorname{ctg} x$	$-1 - \operatorname{ctg}^2 x$	$\arccos x$	$-\frac{1}{\sqrt{1-x^2}}$

A differenciálás szabályait az alábbi táblázat tartalmazza:

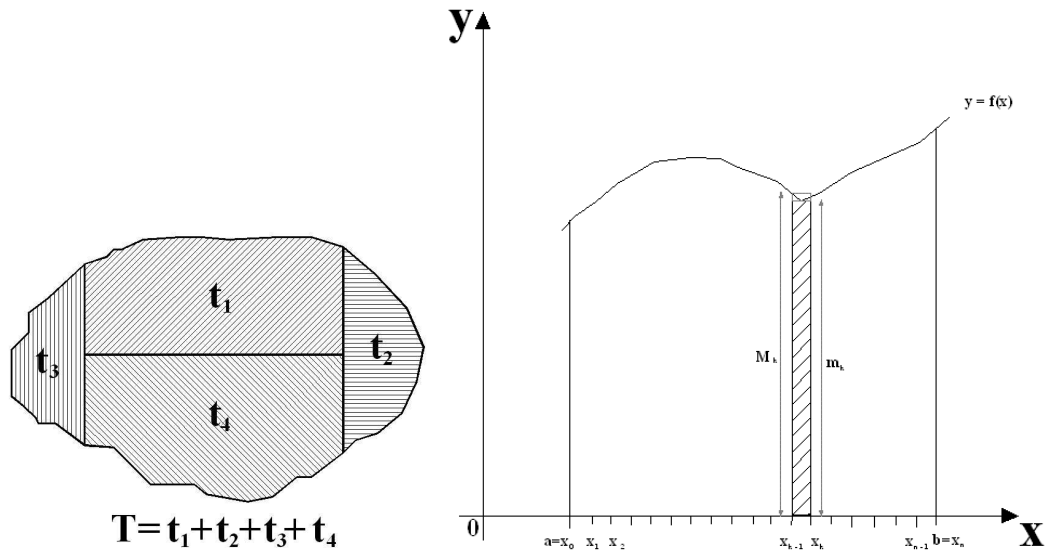
szabály	függvény $f(x)$	differenciálhányados $f'(x)$
összeg	$u(x)+v(x)$	$u'(x)+v'(x)$
szorzás	$u(x) \cdot v(x)$	$u'(x) \cdot v(x) + u(x) \cdot v'(x)$
állandóval szorzás	$c \cdot u(x)$	$c \cdot u'(x)$
hányados	$\frac{u(x)}{v(x)}$	$\frac{u'(x) \cdot v(x) - u(x) \cdot v'(x)}{v^2(x)}$
Összetett függvények	$u[v(x)]$	$u'[v(x)] \cdot v'(x)$

### 3.2. Integrálszámítás

Az integrálszámítás fogalmához két olyan probléma vezet, amelyek első pillantásra semmi összefüggésben nincsenek egymással. Az egyik problémafelvetés az, hogy hogyan lehet egy görbe vonallal határolt síkrész területét kiszámítani, a másik pedig, hogy valamely  $y = f(x)$  függvényhez hogyan lehet meghatározni egy olyan  $F(x)$ , úgynevezett primitív függvényt, amelynek a differenciálhányadosa  $f(x)$ ,  $F'(x) = f(x)$ . Ez a differenciálás fordított művelete.

Az egyszerű geometriai alakzatok (négyzet, téglalap, háromszög, trapéz, kör, stb.) területének kiszámítására vannak képleteink, de vajon hogyan lehetne kiszámítani görbe vonalakkal határolt területeket (pl. tó, rét, erdő, stb.). A teljesen zárt görbével határolt területeket feldarabolhatjuk úgy, hogy a kapott részterületeknek csak az egyik oldala lesz görbe, amely a görbe vonalú határ egy darabja.

Ezért matematikailag elegendő azzal a kérdéssel foglalkozni, hogy hogyan határozhatunk meg egy olyan görbe  $[f(x)]$  alatti területet, amelynek másik három oldalát az  $x$  tengely egy  $[a, b]$  intervalluma, továbbá az  $a$  és  $b$  pontokban a függvényhez húzott,  $y$  tengellyel párhuzamos szakaszok határolják.



Ennek a területét közelítőleg úgy számíthatjuk ki, hogy az  $a \leq x \leq b$  intervallumot  $n$  egyenlő részintervallumra ( $a = x_0 \leq x \leq x_1$ ;  $x_1 \leq x \leq x_2$ ; ...;  $x_{k-1} \leq x \leq x_k$ ; ...;  $x_{n-1} \leq x \leq x_n = b$ ) osztjuk. Jelöljük az  $[x_0, x_1]$  intervallumon felvett legkisebb függvényértéket  $m_1$ -gyel, a legnagyobbat pedig  $M_1$ -gyel, az  $[x_1, x_2]$  intervallumon felvett legkisebb függvényértéket  $m_2$ -vel, a legnagyobbat pedig  $M_2$ -vel, az  $[x_{k-1}, x_k]$  intervallumon felvett legkisebb függvényértéket  $m_k$ -val, a legnagyobbat pedig  $M_k$ -val, végül az  $[x_{n-1}, x_n]$  intervallumon felvett legkisebb függvényértéket  $m_n$ -nel, a legnagyobbat pedig  $M_n$ -nel. Ha a függvény az  $[a, b]$  intervallum minden pontjában



ban folytonos, akkor a részintervallumokon is folytonos és minden részintervallumra igaz, hogy létezik legalább egy olyan hely, ahol a függvény a legkisebb és legalább egy hely, ahol a legnagyobb értékét veszi fel.

Rajzoljunk minden intervallum fölé két olyan téglalapot, amelyek közül az egyik magassága a legkisebb ( $m_1, m_2, \dots, m_n$ ), a másiké pedig a legnagyobb ( $M_1, M_2, \dots, M_n$ ) függvényértékkel egyenlő. Számítsuk ki az  $f(x)$  függvénygörbe alatti síkidomot részben lefedő, úgy is mondhatnánk, hogy beírt téglalapok területének összegét. Ezt jelöljük  $t_n$ -el.

$$t_n = m_1(x_1 - x_0) + m_2(x_2 - x_1) + \dots + m_k(x_{k-1} - x_k) + \dots + m_n(x_{n-1} - x_n)$$

Számoljuk ki az  $f(x)$  függvénygörbe alatti síkidomot teljesen lefedő és a síkidom részben túlnyúló (körülírt) téglalapok területének összegét. Ezt jelöljük  $T_n$ -el.

$$T_n = M_1(x_1 - x_0) + M_2(x_2 - x_1) + \dots + M_k(x_{k-1} - x_k) + \dots + M_n(x_{n-1} - x_n)$$

Megállapíthatjuk, hogy a beírt téglalapok területének összege kisebb, a körülírt téglalapok területének összege pedig nagyobb, mint a függvény alatti terület ( $t_n < T < T_n$ ). Minél több részre osztjuk az  $[a, b]$  intervallumot, annál jobban megközelítjük a függvény alatti területet.

A  $t_1, t_2, \dots, t_n$  számok sorozatát az alsó összegek sorozatának, a  $T_1, T_2, \dots, T_n$  számok sorozatát pedig a felső összegek sorozatának szokták nevezni, amelyeknek létezik közös határértéke:

$$\lim_{n \rightarrow \infty} t_n = \lim_{n \rightarrow \infty} T_n = A$$

Az  $A$  az olyan egyetlen olyan szám, amely minden  $n$ -re  $t_n$  és  $T_n$  közé esik ( $t_n < T_n$ ). Nyilvánvaló, hogy az  $A$  határérték a keresett  $T$  terület, amely az  $a \leq x \leq b$  intervallum fölött van. Ezt a területet az  $y = f(x)$  függvény Riemann szerinti értelemben vett integráljának nevezzük. Jele:

$$A = \int_a^b f(x) dx \quad (\text{ejtsd.: integrál a-tól b-ig } f(x) dx)$$

A határozott integrál segítségével megoldható a bevezetőben már említett másik probléma, tehát egy folytonos  $f(x)$  függvényhez olyan  $F(x)$  primitív függvényt találunk, amelyre  $F'(x) = f(x)$ . Ha egy határozott integrál felső határát megváltoztatjuk, akkor más értéket kapunk. Ezt könnyű belátni, ha egy pozitív  $y = f(x)$  függvény határozott integráljára, mint területre gondolunk. Ebből következik, hogy az  $F(b)$  határozott integrál függvénye a felső határának,  $b$ -nek. Mivel a független változót  $x$ -el szokás jelölni, ezt így írhatjuk:

$$F(x) = \int_a^x f(t) dt$$

Ez az  $y = F(x)$  függvény az  $y = f(x)$ -nek egy primitív függvénye, mert ha az  $f(x)$  folytonos az  $a \leq t \leq b$  intervallumban, akkor  $F(x)$  az  $a \leq x \leq b$  intervallumban differenciálható, és  $F'(x) = f(x)$ . Ezen a primitív függvényen kívül még végtelen sok további is van, amelyek ettől csak egy pozitív összeadandóval különböznek. Két primitív függvény különbsége mindig egy  $x$ -től független konstans. Ezért az  $f(x)$  primitív függvényét szokták még  $F(x) + c$  alakban is felírni, ahol a  $c$  egy konstans. Geometriailag ez azt jelenti, hogy az összes primitív függvény egymás eltoltja az  $y$  tengely irányában.

Valamely  $f(x)$  függvény minden primitív  $F(x)$  függvényét az  $f(x)$  határozatlan integráljának nevezzük és így jelöljük:

$$F(x) = \int f(x)dx$$

A határozatlan integrál kiszámítása nem könnyű feladat, de a differenciálszámításból sok függvény deriváltja ismert és azok alapján összeállított táblázat jól használható.

Legyen az  $F(x) = \int f(x)dx$  primitív függvény, a  $k$  és  $c$  állandók, továbbá  $n$  egész szám ( $n \neq -1$ ), akkor az alábbi táblázatban az  $f(x)$  primitív függvényei  $[F(x)]$  megtalálhatók.

$f(x)$	$F(x)$	$f(x)$	$F(x)$
$k$	$k \cdot x + c$	$\cos x$	$\sin x + c$
$x^n$	$\frac{x^{n+1}}{n+1} + c$	$\sin x$	$-\cos x + c$
$\frac{1}{x}$	$\ln x  + c$	$\frac{1}{\cos^2 x}$	$\operatorname{tg} x + c$
$a^x$	$\frac{a^x}{\ln a} + c$	$\frac{1}{\sin^2 x}$	$-\operatorname{ctg} x + c$
$e^x$	$e^x + c$	$\cosh x$	$\sinh x + c$
$\ln x$	$x \cdot \ln x - x + c$	$\sinh x$	$\cosh x + c$

Valamely  $\int_a^b f(x)dx$  határozott integrálnak a kiszámítása a gyakorlatban

csak nagyon egyszerű függvényeknél végezhető el, ezért a határozott és a határozatlan integrál közötti, most tárgyalt összefüggés segíthet. Ha  $F(x)$  az  $f(x)$  határozatlan integrálja és ez a táblázatban megtalálható, akkor az

$\int_a^b f(x)dx = F(b) - F(a)$  egyenlet alapján elegendő a primitív függvénynek csak

az alsó és a felső határán felvett értékeit kiszámítani és  $F(a)$ -t kivonni az  $F(b)$ -ből.

A fentiek alapján belátható, hogy a határozott integrál nem negatív és folytonos függvények esetén a függvénygörbe alatti terület mérőszámát adja.

**Tulajdonságai:**

- Az integrál értékének kiszámításából következik, hogy  $\int_a^a f(x)dx = 0$ .
- Az  $f(x)$  és  $g(x)$  folytonos függvényekről feltételezzük, hogy  $f(x) \geq 0$ ,  $g(x) \geq 0$ ,  $f(x) - g(x) \geq 0$ . Ebből következik, hogy  $f(x) + g(x) \geq 0$ , továbbá  $f(x) + g(x)$  függvény is folytonos. A folytonosság miatt ezek integrálja is létezik.

$$\int_a^b [f(x) \pm g(x)dx] = \int_a^b f(x)dx \pm \int_a^b g(x)dx$$

- Ha az  $f(x) \geq 0$  folytonos függvény és  $a < c < b$ , akkor az  $f(x)$  függvényre igaz, hogy

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx$$

- Ha az  $f(x)$  függvény az  $[a, b]$  intervallumon integrálható, akkor ugyanezen az intervallumon a  $c$  állandóval szorzott  $c \cdot f(x)$  függvény is integrálható, és

$$\int_a^b c \cdot f(x)dx = c \cdot \int_a^b f(x)dx$$

**Példa:**

1, Egy téglalap alakú csatornában (szélesség 4 m, mélység 2 m) sebességmérést végzünk. Adatainkra jól illeszthető a  $v(x, z) = 0,3z^2(1 - \frac{|x|}{2})$  függvény, ahol  $x$  a csatorna közepétől mért vízszintes távolság méterben,  $z$  a csatorna aljától mért magasság méterben, míg  $v$  a felületre merőleges sebesség m/s-ban van megadva. Számoljuk ki a vízhozamot és az átlagsebességet!

Megoldás:

Mivel a kevés számú adatra jól illeszkedő függvénykapcsolatot találtunk, a vízhozam összegzés helyett integrálással is kiszámítható, az integrandus  $dA = dx dz$ . Mivel a feladat szimmetrikus, így elég a  $0 \leq x \leq 2$ ;  $0 \leq z \leq 2$  között integrálni. A  $dA$  felületen  $dQ = v(x, z)dA$  nagyságú vízhozammal áramlik át a víz.

$$\begin{aligned}
Q &= \int_A dQ = \int_A v(x, z) dA = \int_0^2 \int_{-2}^2 v(x, z) dx dz = 2 \int_0^2 \int_0^2 v(x, z) dx dz = 2 \int_0^2 \int_0^2 0,3z^2 \left(1 - \frac{x}{2}\right) dx dz = \\
&= 2 \int_0^2 0,3z^2 \int_0^2 \left(1 - \frac{x}{2}\right) dx dz = 2 \int_0^2 0,3z^2 \left[ x - \frac{x^2}{4} \right]_0^2 dz = 2 \int_0^2 0,3z^2 \left( 2 - \frac{2^2}{4} - 0 + \frac{0^2}{4} \right) dz = \\
&= 0,6 \int_0^2 z^2 dz = 0,6 \left[ \frac{z^3}{3} \right]_0^2 = 0,6 \left( \frac{2^3}{3} - \frac{0^3}{3} \right) = 1,6
\end{aligned}$$

Azaz a vízhozam  $1,6 \text{ m}^3/\text{s}$ , a középsebesség:

$$\bar{v} = \frac{Q}{A} = \frac{1,6 \frac{\text{m}^3}{\text{s}}}{4 \text{ m} \cdot 2 \text{ m}} = 0,2 \frac{\text{m}}{\text{s}}$$

2, Hasonlítsuk össze ezt az értéket a súlyozott átlag módszerével kapott értékekkel, ha a mért adatok a következők:

$ x $ (m)	$z$ (m)	$v(x, z)$ (m/s)
0,5	0,5	0,05625
0,5	1,5	0,50625
1,5	0,5	0,01875
1,5	1,5	0,16875

Megoldás: Ezek a keresztmetszvény  $1 \cdot 1$ -es felületelemeinek középpontjai, így

$$\bar{v} = \frac{2 \cdot (0,05625 + 0,050625 + 0,01875 + 0,16875)}{8} = 0,1875 \frac{\text{m}}{\text{s}}$$

Azaz ebben az esetben a különbség  $6,25 \%$ .

### Feladatok

- 1, Egy lejtőt az  $f(x) = -(x-3)^2 + 9$  függvény ír le, ahol  $x$  a lejtő aljától mért lejtőtalap méterben,  $0 \leq x \leq 6$ ,  $f(x)$  pedig a lejtő aljától mért magasság szintén méterben. Milyen meredek a lejtő az  $x = 6$ -nál? Mennyi ott a magassága? Mennyi anyag nyerhető ki belőle ( $x = 6$ -nál)?
- 2, Határozzuk meg az előző feladatban megadott  $f(x)$  függvény alatti területet (a lejtő keresztmetszetét) a  $[0, 6]$  intervallumban!
- 3, Mekkora a területe egy szabályos ellipszis alakú szélbarázdának, ha a hosszúsága  $80 \text{ m}$ , szélessége  $30 \text{ m}$ ?
- 4, Határozza meg az  $1000 \text{ m}$  vastag tengervízoszlop átlagos sűrűségét, ha a tengervíz sűrűsége  $\rho = \rho_0 \cdot (1 + 0,0001h)$  képlet szerint változik a mélységgel ( $h$ )?  $\rho_0 = 1030 \text{ kg/m}^3$

## 4. Földi pontok koordinátáinak átszámítási lehetőségei. Vetületi transzformációk. Mérések, számítások pontosságának vizsgálata.

### 4. 1. Transzformációk

Azokat a függvényeket, amelyeknek értelmezési tartománya is, értékkészlete is ponthalmaz, **geometriai transzformációknak** nevezzük.

A geometriai transzformációk közül elsősorban azokkal foglalkozunk, amelyek a geoinformatikai (GIS) módszerekkel történő geográfiai feldolgozások során előfordulhatnak. A cél a szükséges fogalmak meghatározása, a fontosabbnak vélt tételek ismertetése (a matematikai összefüggések igazolása nélkül) és szakmai példák megoldása, gyakorlása.

A koordinátarendszerek transzformációjánál – rövidebben: koordináta-transzformációnál – a tér pontjai helyben maradnak és a koordináta-rendszert transzformáljuk hozzájuk képest. Ilyenkor megadjuk azt a **geometriai transzformációt**, amely az eredeti koordináta-rendszert – kezdőpontját, tengelyeit és egységpontját – átvizsi az újba és a pontoknak az új rendszerbeli koordinátáit ki kell számolni. (Ha például a rendszert dx-szel eltoljuk, a pontok új x koordinátája dx-szel változik.) A koordináta-transzformáció gyakori esete az áttérés új koordináta-rendszerre. Ezen kívül beszélhetünk egy térnek egy másik térre, az első egy koordináta-rendszeréből a másik egy koordináta-rendszerébe való leképezésről.

Abban az esetben, ha egy ponthoz a transzformáció egy pontot rendel, azt mondjuk, hogy ezek egymásnak megfelelő pontok. Tulajdonképpen egy alakzathoz hozzárendeljük a pontjaihoz tartozó képpontokból álló alakzatot. Így az alakzat transzformálása a képalakzatra való áttérést jelenti.

Kölcsönösen egyértelmű leképezésről akkor beszélünk, ha más-más tárgyponthoz más-más képpontot rendelünk. A sík vagy tér transzformációja elfajuló, ha a teljes sík képe lineáris alakzat, illetve a teljes tér képe síkbeli alakzat lesz (pl. térkép). A kölcsönösen egyértelmű transzformációnak van ellentétes (inverz) transzformációja, amely a képponthoz rendeli a tárgypontokat.

A geometriai transzformációk az alábbi tulajdonságokkal jellemezhetők:

- **egyenestartó:** ha egy egyenes képe a transzformáció után is egy egyenes.
- **távolságtartó:** ha a transzformáció után a tárgypontok távolsága megegyezik a képpontok távolságával.

Hasonlóan értelmezhetők a szakasztartó, szögtartó, párhuzamosságtartó, területtartó transzformációk is.

- A geometriai transzformáció **fixpontjának** nevezzük az olyan pontot, amely önmagának a képe.
- Fix egyenesnek nevezzük az olyan egyenest, amelynek minden pontja fixpont.

- Több transzformáció egymás utáni alkalmazásaként nyert transzformációt a **transzformációk szorzatának** nevezzük.
- Az olyan transzformációt, amely minden pontot helyben hagy, **identikus leképezésnek** nevezzük.

A legegyszerűbb geometriai transzformációk az **egybevágóságok**. Az egybevágóságnak van inverze és az is egybevágóság. A távolságtartó leképezést egybevágóságnak nevezzük. Az egybevágóságok speciális esetei a tengelyes tükrözés, a pont körüli elforgatás, a középpontos tükrözés és az eltolás. A pont körüli elforgatás, a középpontos tükrözés és az eltolás szemléletesen olyan, mintha az egész tartalmazó síkot (vagy teret) elmozgatnánk az alakzattal együtt.

A transzformáció neve	Fix pontja(i)	Helyettesíthető
egyenesre (tengelyre) vonatkozó tükrözés	a tengely minden pontja	a tengely körüli 180°-os (térbeli) elforgatás
pontra vonatkozó tükrözés	a tükrözés középpontja	a tükrözés középpontján áthaladó és egymásra merőleges két egyenesre történő egymás utáni tükrözéssel
Pont körüli $0^\circ \leq \alpha < 360^\circ$ nagyságú adott irányú elforgatás	az elforgatás középpontja, $0^\circ = \alpha$ esetén minden pont	az elforgatás középpontján áthaladó $t_1$ és $t_2$ egyenesekre történő egymás utáni tükrözéssel. A $t_1$ egyenes $2\alpha$ forgásszöggel vihető át a $t_2$ egyenesbe
$d \geq 0$ nagyságú, adott irányú eltolás	$d \neq 0$ esetén nincs, $d = 0$ esetén minden pont	az eltolás irányára merőleges és egymástól $d/2$ távolságban lévő párhuzamos egyenesekre történő egymás utáni tükrözéssel (a tükrözések sorrendje lényeges!)

**Hasonlósági transzformációnak** nevezünk egy pontranszformációt akkor, ha bármely két képpont távolságát osztva a tárgyponatok távolságával, ugyanazt a pozitív valós számot kapjuk. Tehát, ha a transzformáció A-hoz A'-t, B-hez B'-t rendeli, akkor  $dA'B'/dAB=1$  minden A, B pont esetén.

A **középpontos hasonlóság** szintén rendelkezik a definícióban előírt tulajdonsággal, tehát hasonlósági transzformáció. Minden hasonlóság megkapható, mint egy egybevágóság és egy középpontos hasonlóság szorzata. Jelöljünk ki egy O pontot, és adjunk meg egy  $\lambda > 0$  valós számot. Az O ponthoz rendeljük hozzá önmagát. Egy tetszőleges, de O-tól különböző P ponthoz rendeljük hozzá azt az OP félegyenesre eső P' pontot, amelyre  $dOP' = \lambda dOP$ . Az így definiált pontranszformációt középpontos hasonlóságnak nevezzük.

*kicsinyítés, ha  $\lambda < 1$*   
*A hasonlóság egybevágóság, ha  $\lambda = 1$*   
*nagyítás, ha  $\lambda > 1$ .*

Két alakzat akkor hasonló, ha van olyan hasonlósági transzformáció, amely az egyiket a másikhoz rendeli. A hasonlósági transzformáció fontos tulajdonsága a szögtartás. Ebből következik az is, hogy az egy egyenesre eső pontoknak ugyanolyan pontok felelnek meg, tehát a hasonlóság egyenestartó transzformáció.

Körök hasonlóságára vonatkozóan megállapítható, hogy két kör hasonló, mert az egyik kör középpontjához a másik kör középpontját rendeli, s a két kör sugarának aránya a hasonlóság aránya, amely a két kört egymásba viszi.

## 4. 2. Koordináta transzformációk

Abban az esetben, ha egy terület objektumai különböző koordináta rendszerben vannak (pl. különböző vetületi rendszerben lévő térképeken), akkor a meglévő elemeket úgy kell transzformálni, hogy közös koordináta rendszerbe kerüljenek.

A koordináta-transzformációt tulajdonképpen a következő módon hajthatjuk végre:

- az objektum koordinátáinak megváltoztatása érdekében mozgatjuk az objektumot a fix koordináta rendszerben, vagy
- az objektumot tartjuk fixen és elmozgatjuk a koordináta rendszert.

Mindegyik esetben az objektum „régii” helyét meghatározó koordináta értékek  $(x, y)$  az új koordináta rendszerben megváltoznak  $(u, v)$ .

A koordináta transzformáció lehet **affin** (eltolás, méretarányváltozás, forgatás, tükrözés), vagy nem lineáris, ekkor **magasabbrendű** transzformációról beszélünk.

Akkor nevezzük affinnak a transzformációt, ha a párhuzamos vonalak párhuzamosak maradnak. A magasabbrendű transzformációnál nem szükséges, hogy a vonalak egyenesek és párhuzamosak maradjanak.

Az affin transzformációnak a következő négy típusát különböztetjük meg:

**a, Eltolás** (ha a koordinátarendszer kezdőpontja mozog, de a tengelyek nem fordulnak el)

$$P(x, y) \quad P'(u, v) \quad (u = x - \mathbf{a}; \quad v = y - \mathbf{b})$$

A rendszer kezdőpontja '**a**' egységgel mozdult el az x tengellyel párhuzamosan, és '**b**' egységgel az y tengellyel párhuzamosan.

**b, Méretarányváltozás** (ha a koordináta-rendszer kezdőpontja és a tengelyek változatlanok, de a méretarány változik)

$$P(x, y) \quad P'(u, v) \quad (u = cx; \quad v = dy)$$

Így az  $x$  és  $y$  méretaránytényezői különbözőek és az objektum alakja is megváltozik.

**c, Forgatás** (ha a koordináta-rendszer kezdőpontja fix, de a tengelyek a kezdőpont körül elfordulnak)

$$P(x, y) \quad P'(u, v) \quad \begin{cases} u = x \cos(\alpha) + y \sin(\alpha); \\ v = -x \sin(\alpha) + y \cos(\alpha) \end{cases}$$

Az  $\alpha$  szöveget az óramutató járásával ellentétes irányban mérjük!

**d, Tükrözés** (a koordináta-rendszer fordított, az objektum képe az eredeti tükörképe lesz)

$$P(x, y) \quad P'(u, v) \quad (u = x; \quad v = -y)$$

A gyakorlati feladatoknál előfordulhat, hogy az objektumot csak több affin transzformációval tudjuk beilleszteni az új koordináta rendszerbe (pl. eltolás, forgatás és méretarány változtatásával). Ilyenkor ügyelni kell a sorrendre, mert a hatás különböző lehet (pl. egy méretarány változtatás utáni eltolás nem ugyanaz, mint egy eltolás és utána méretarány változtatás).

Az összetett affin transzformációnál az új koordináták meghatározása is bonyolultabb pl.:

$$P(x, y) \quad P'(u, v) \quad (u = a + bx + cy; \quad v = d + ex + fy)$$

### 4.3. Magasabbrendű transzformációk

A *nem lineáris transzformációknál* az új koordináták kiszámításánál olyan egyenleteket használunk, amelyek görbe felületeket írnak le. Ilyenkor a pontokat nem egyenletesen transzformáljuk, az is előfordulhat, hogy a párhuzamos egyenesek a transzformáció után görbékké alakulnak.

Lehetőség van arra is, hogy a térképet részekre bontjuk és külön-külön végezzük el a transzformációt. Ekkor minden résznek más lesz a transzformációs egyenlete. Fontos, hogy az egyenleteknek ki kell elégíteniük az élek folytonosságának feltételeit a részek csatlakozásánál.

A magasabbrendű transzformációk általában nagyobb pontosságot biztosítanak. A pontosság alatt azt értjük, hogy a transzformációt az illesztőpontokra alkalmazva, hűen visszacapjuk azok másik koordináta-rendszerbeli koordinátáit.



## *Geográfiai alkalmazások*

A földrajzi alkalmazások során a transzformálható objektumok általában valamilyen földrajzi koordinátákkal rendelkeznek, ezért elsősorban koordináta-transzformációk elvégzésére van szükség. Természetesen olyan feladat is adódhat, hogy a térképlap vetülete nem ismert, vagy nem pontos, ezért először a keretjeleket (ha van) és pontokat azonosítani kell.

A koordináta-rendszer kiválasztása után a rendszerben elhelyezett és koordinátákkal azonosított objektumokon már a fenti módszerekkel végrehajthatjuk a kívánt transzformációt. A transzformáció számítógépen elvégezhető. A fontosabb transzformációs egyenleteket a szoftverekbe beépítették.

A geoinformatikai alkalmazásoknál a transzformáció előtt meg kell vizsgálnunk az állományok formátumát, ugyanis a raszteres térképek, úrfelvételek, a digitalizált vagy a geodéziai szoftverrel készített formátumok általában nem egyeznek meg a GIS szoftver saját formátumaival. Ahhoz tehát, hogy a GIS tárolni tudja az importált adatokat először azok formátumát kell transzformálni. A formátum transzformáció eredményességétől függ a GIS használhatósága.

A raszteres adatoknál fontos a „kép” sor és oszlop méretének és a pixel attribútum értékeknek az ismerete. Ha a GIS transzformációs program felismeri ezeket, akkor nincs akadálya a formátum transzformációnak. Ha ez nem sikerül, akkor a kép értékeit is át kell alakítani a szoftver formátumban alkalmazott értékekre.

A vektoros állományoknál még bonyolultabb lehet a formátum transzformáció. A vektoros digitális térképek adatállománya különböző modellben kerül tárolásra. Gyakori a spagetti, illetve topológiai modell alkalmazása. A spagetti állományt gyakran alkalmazzák az automatikus térképező szoftverek, míg a GIS rendszerint topológiai modellt használ. A térképezésre gyakran használt általános rajzoló szoftver az AUTOCAD pedig a létrehozó utasítástól függően mindkét modellben létrehozhatja az alakzatokat.

A transzformáció csak akkor lesz eredményes, ha az általunk használt GIS szoftver felismeri az importált formátum adatmodelljét és azt átalakítja a saját adatmodelljére. Ez a feladat egyértelműen megoldható, ha mind az exportáló mind az importáló rendszer egységes adatmodellel rendelkezik. Nem ilyen egyszerű azonban a helyzet az AUTOCAD esetében, ugyanis a gyakorlati tapasztalatok azt bizonyítják, hogy nem transzformálhatók eredményesen az IDRISI szoftverbe azok az AUTOCAD dxf formátumú poligonok, melyek 'line' utasítással készültek, vagy gyakorlatilag használhatatlanok az AUTOCAD-ban azok a dxf fájlok, amelyeket az ITR nevű automatikus térképező programból visznek át. Ez utóbbi program ugyanis spagetti modellű és az átvitt sokszögek az AUTOCAD-ban önálló vonalak együtteseiként jelennek meg.

A geodéziai-geometriai transzformációk megoldása sem egyszerű. A térbeli adatok rendszerint valamely sík vetületi koordináta rendszer felhasználásával készültek. Ritkán fordul elő, hogy az adatfájlok koordinátái ugyanarra a vetületi rendszerre vonatkoznak. Magyarországon a különböző időszakokban készült térképek vetületi rendszere eltérő, ezért különféle transzformációk alkalmazására van szükség akkor, ha ugyanarról a területről eltérő vetületű térképeket akarunk használni.

A feladat összetett és a pontos munka nagy odafigyelést igényel. A papíralapú térképeket digitálisra kell átalakítani. A GIS feldolgozásokhoz gyakran vektoros állományokat kell előállítani. Ez különböző módszerekkel történhet.

- Hagyományos módszerrel a térképen lévő objektumok adatait digitalizálással végezzük. A térképeket digitalizálhatjuk digitalizáló tábla segítségével. Ennél a módszernél a térkép tartalma a digitalizáló tábla koordinátaiban kerül rögzítésre, mégpedig a táblára felerősített térkép tényleges tájolásának megfelelően. Itt arra is ügyelni kell, hogy a digitalizálás befejezése előtt a térkép helyzetét a táblán ne változtassuk meg, vagy minden elmozdításnál meg kell határoznunk illeszkedési pontokat, amelyek segítségével az egyes részek összeillesztése (transzformációval) elvégezhető.
- A digitalizálást gyakran, valamelyik szoftver felhasználásával, a képernyőn végezzük. Ilyenkor először egy raszteres állományt kell készítenünk a térkép beszkenelésével. Ez újabb problémát vet fel, és sok hiba elkövetésére ad lehetőséget. (pl. a részletekben beszkenelt térképrészek összeillesztése, felbontás kiválasztása, torzulások, stb.) A képernyőn történő digitalizálásnál a térképi objektumok adatai általában a szoftver koordináta rendszerében tárolódnak. Az a szerencsésebb eset, ha a raszteres állományt először transzformáljuk a kívánt koordináta rendszerbe, és csak ezután kezdjük el a digitalizálást.
- A raszteres állomány digitalizálása történhet „automatikusan” is. Erre több szoftvergyártó cég is fejlesztett ki szoftvereket. Topográfiai térképeknél ezek használata nagyon körülményes és nagy odafigyelést igényel (pl. a szintvonalak és az utak metszéspontjánál a szoftver nem tudja eldönteni a haladási irányt és a kezelő utasítására vár.).

A digitalizálással elkészített, ellenőrzött (javított) különböző vetületi rendszerű rétegeket csak úgy tudjuk közösen feldolgozni, ha egy közös, rendszerint a hivatalos (házánkban az EOVS) koordináta rendszerbe transzformáljuk.

Két különböző vetületi rendszerben készített térképet csak akkor tudunk pontos képletekkel közös rendszerbe transzformálni, ha a kérdéses vetületi rendszerek ugyanarra a dátumra vonatkoznak. A hazai térképeknél azonban gyakran

előfordul, hogy a vetületi rendszerek különböző időpontúak, ezért a transzformációnál erre is oda kell figyelni.

A GIS szoftverek tartalmaznak (napjainkban már hazánkra is érvényes) transzformációs eljárásokat és ez megkönnyíti a térképek vetületbe illesztését. Azonban, ha nem találjuk a megfelelő transzformációt, akkor a feladatot ismert pontok koordinátáinak felhasználásával, a „gumilepedő” transzformációval tudjuk megoldani. Ezt valóban úgy kell elképzelnünk, mintha a transzformálandó térkép egy gumilepedőn volna, amelyet ráfeszítünk egy kiválasztott koordináta rendszerre, illetve az adott koordináta-rendszerben készült térképre. Erre azonban csak akkor kerülhet sor, ha a két különböző vetületű térképen találunk megfelelő számú közös pontot. Arra ügyeljünk, hogy a térkép koordináta-hálózati metszéspontjai erre a célra nem használhatók. A fokhálózati vonalak metszéspontjai is csak akkor használhatók, ha a két vetület dátuma megegyezik. Megfelelő számú és jó területi eloszlású, ismert koordinátájú geodéziai alappontok azonban jelentősen segítik és pontosabbá teszik a transzformációt.

#### 4. 4. Hazai térképvetületek transzformációja

Gyakori feladat a felszíni pontok koordinátáinak meghatározása. Napjainkban erre legalkalmasabb a GPS technika. A jobb GPS műszereknél lehetőség van a vetületi rendszer beállítására. Abban az esetben, ha az alkalmazott vetületek tulajdonságait nem ismerjük, akkor a mérés eredményeinek megadása után jelentős hibákat követhetünk el (pl. ugyanannak a terepi pontnak a földrajzi koordinátái a különböző vetületi rendszerekhez tartozó alapfelületeken mások és mások).

Az elmúlt századokban hazánkban különféle alapfelületeket használtak és eltérőek a különböző időszakban készült térképek vetületi rendszere is. A transzformációhoz ismernünk kell a vetületek főbb jellemzőit. A legfontosabb vetületekkel a térképtan tananyagban foglalkoztunk, ezért most csak a transzformációhoz nélkülözhetetlen legfontosabb adatokat\* foglaljuk össze.

A második katonai felvételnél használt ellipszoid fél nagytengelye (a) 3362035 öl, lapultsága ( $\alpha$ ) 1/310. A hálózat csillagászati kezdőpontja a bécsi Szent István székesegyház (Stephansdom) tornya volt. Az ábrázolásnál fellépő eltérések miatt a Monarchia területén több koordináta-rendszert vezettek be. A bevezetett tíz (ún. *vetületnélküli*) rendszer közül a Magyar Királyság területére az alábbi három esett:

*a, Budai rendszer.* Kezdőpontja a volt gellérthegyi csillagda keleti pillére, amelynek a bécsi Szent István székesegyház tornyából levezetett alapfelületi koordinátái:

$$\varphi = 47^{\circ} 29' 14,97''$$

---

\* A vetületek leírása a [http://www.agt.bme.hu/staff\\_h/varga/vetulet.html](http://www.agt.bme.hu/staff_h/varga/vetulet.html) alapján készült.

$$\lambda = 36^{\circ} 42' 51,57'' \quad (\text{Ferrotól számítva})$$

*b, Nagyszebeni rendszer.* Kezdőpontja a *Vízakna (Vízaknai-hegy)* nevű háromszögelési pont, melynek alapfelületi koordinátáit a *budai rendszer* kezdőpontjából vezették le:

$$\varphi = 45^{\circ} 50' 25,13''$$

$$\lambda = 41^{\circ} 46' 32,71'' \quad (\text{Ferrotól számítva})$$

*c, Ivanic-i (továbbiakban: ivanicsi) rendszer.* Kezdőpontja az *Ivanics zár-datorony* háromszögelési pont, melynek csillagászatilag meghatározott alapfelületi koordinátái:

$$\varphi = 45^{\circ} 44' 21,25''$$

$$\lambda = 34^{\circ} 05' 09,16'' \quad (\text{Ferrotól számítva})$$

A fenti három síkkoordináta-rendszer délnyugati tájolásúak, tehát a kezdőmeridián az *x* tengely pozitív ága déli, az *y* tengely pozitív ága pedig nyugat felé mutat.

A harmadik katonai felmérés háromszögelési alapja lényegében azonos a második katonai felméréshez készült háromszögelési hálózattal, de a pontok koordinátáit a régi mérési adatokból összeállított láncolatok segítségével újra számították. Szabatos módszerrel azonban ezt sem egyenlítették ki. Ekkor készültek azok az 1:75 000 és 1:200 000 méretarányú topográfiai térképek, amelyek lapjait az 1950-es évekig - a Gauss- Krüger vetületű topográfiai térképek megjelenéséig használták.

Az I. világháború végéig a földrajzi hosszúságokat a ferroi meridiántól mint kezdőmeridiántól mérték, később a topográfiai térképeken megjelentek a greenwichi meridiántól mért hosszúsági értékek.

A ferroi és greenwichi hosszúságok különbsége:

$$\lambda_F - \lambda_G = 17^{\circ} 39' 46''$$

A magyarországi sztereografikus vetületi rendszerek a következők:

*a, Budapesti rendszer.* Kezdőpontja a *Gellért-hegy nevű háromszögelési pont* Gauss-gömbi megfelelője. Ennek 1908-ig használt földrajzi koordinátái a Bessel-ellipszoidon:

$$\varphi_0 = 47^{\circ} 29' 09,6380''$$

$$\lambda_0 = 36^{\circ} 42' 53,5733'' \quad (\text{Ferrotól számítva})$$

és a régi Gauss-gömbön

$$\varphi_0 = 47^{\circ} 26' 21,1372''$$

$$\lambda_0 = 0^{\circ} 0' 0,00000''$$

b, *Marosvásárhelyi rendszer*. Kezdőpontja *Kesztej-hegy* nevű háromszögelési pont Gauss-gömbi megfelelője, melynek földrajzi koordinátái a Bessel-ellipszoidon:

$$\varphi_0 = 46^\circ 33' 06,4273''$$

$$\lambda_0 = 42^\circ 03' 20,9550'' \text{ (Ferrotól számítva)}$$

Mindkét rendszerben a kezdőpont meridiánjának egyenesként ábrázolt képe a síkkoordináta-rendszer *x* tengelye, a kezdőpontban a meridiánra merőleges gömbi főkör (ortodroma) szintén egyenes képe pedig az *y* tengely. Az *x* tengely pozitív ága délre, az *y* tengelyé nyugatra mutat (DNY-i tájékozású koordináta-rendszerek).

Az *Állami Földmérés* és 1937-ig a *katonai topográfia* is az előbbi koordináta-rendszerekben értelmezett  $y_{ST}$  és  $x_{ST}$  koordinátákat alkalmazta. 1937-től a negatív előjelű koordináták kiküszöbölése céljából a topográfiában az  $y_{ST}$ ,  $x_{ST}$  vetületi koordinátákat egy *C* állandóból kivonva nyerték az  $y_{KST}$ ,  $x_{KST}$  *katonai sztereografikus koordinátákat*:

$$y_{KST} = C - y_{ST}$$

$$x_{KST} = C - x_{ST}$$

Ilyen módon a katonai rendszerekben a koordináta-rendszer tájolása is korszerűbbé, ÉK-ivé vált.

A *c* állandó értékei az egyes rendszerekben:

*A Budapesti katonai sztereografikus rendszerben:*

$$C = 500\,000 \text{ m.}$$

*A Marosvásárhelyi katonai sztereografikus rendszerben:*

$$C = 600\,000 \text{ m.}$$

*Az Ivanicsi katonai rendszerben:*

$$C = 400\,000 \text{ m.}$$

Az irányszög- és távolságszámításon kívül mindenféle vetületi számítás csak az eltolás nélküli koordinátákkal végezhető, ezért ilyen számítások előtt vissza kell térni a vetületi koordinátákra:

$$y_{ST} = C - y_{KST}$$

$$x_{ST} = C - x_{KST}$$

A sztereografikus síkkoordinátákból az alábbi egyenletekkel nyerjük a gömbi földrajzi koordinátákat:

$$\sin \varphi = \frac{1}{R^2 + \frac{d^2}{4R}} \left[ \left( R - \frac{d^2}{4R} \right) \sin \varphi_0 - x \cdot \cos \varphi_0 \right]$$

$$\operatorname{ctg} \lambda = -\frac{1}{y} \left[ \left( R - \frac{d^2}{4R} \right) \cos \varphi_0 + x \cdot \sin \varphi_0 \right]$$

ahol

$$d^2 = x^2 + y^2.$$

1908-ban érintő hengervetületeket vezettek be. Ha egy pont budapesti sztereografikus és hengervetületi koordinátáiból kiszámítjuk a gömbi földrajzi koordinátákat, egymástól eltérő eredményeket kapunk. Magyarországot három ferdetengelyű érintő hengervetület fedi. A *hengervetületi északi rendszer (HÉR)* a  $47^\circ 55'$  földrajzi szélességtől északra fekvő területekhez, a *hengervetületi középső rendszer (HKR)* a  $46^\circ 22'$  és a  $47^\circ 55'$  földrajzi szélességekkel határolt sávhoz, a *hengervetületi déli rendszer (HDR)* pedig a  $46^\circ 22'$  földrajzi szélességtől délre levő területekhez tartozik. A koordináta-rendszerek tájolása a sztereografikus vetületekéhez hasonlóan DNY-i.

Ha a régi elhelyezésnek megfelelő alapfelületi koordinátákra van szükségünk, előbb sztereografikus koordinátákat számíthatunk a hengervetületekből, majd azokból a sztereografikus vetületnél ismertetett módon Bessel-ellipszoidi koordinátákat.

Ha valamilyen okból mégis az új elhelyezésnek és tájékozásnak megfelelő Bessel-ellipszoidi koordinátákból kívánunk valamelyik ferdetengelyű, érintő hengervetületen síkkoordinátákat számítani, akkor előbb a *Gauss-gömbi* koordinátákat, majd azokból a kérdéses hengervetületre vonatkozó segédföldrajzi koordinátákat számítunk:

$$\varphi' = \arcsin(\sin \varphi \cdot \cos \varphi_0 - \cos \varphi \cdot \sin \varphi_0 \cdot \cos \lambda)$$

$$\lambda' = \arcsin \frac{\cos \varphi \cdot \sin \lambda}{\cos \varphi'}$$

A hengervetületi síkkoordinátákat az alábbi képletekkel számíthatjuk ki:

$$x = -R \cdot \ln \operatorname{tg} \left( 45^\circ + \frac{\varphi'}{2} \right)$$

$$y = -R \frac{\lambda'}{\rho^\circ}$$

ahol  $R$  a régi Gauss-gömb sugara és

$$\rho^\circ = \frac{180^\circ}{\pi}$$

Ha viszont ferdetengelyű, érintő hengervetületi síkkoordinátákból kívánunk az új elhelyezés és tájolás szerinti Bessel-ellipszoidi koordinátákat számítani, akkor előbb gömbi segédföldrajzi koordinátákat a

$$\varphi' = 2 \cdot \operatorname{arc} \operatorname{tg} e^{\frac{-x}{R}} - 90^\circ \text{ és}$$

$\lambda' = -\frac{y}{R} \rho^\circ$  majd a gömbi földrajzi koordinátákat a

$$\varphi = \arcsin(\sin \varphi' \cdot \cos \varphi_0 + \cos \varphi' \cdot \sin \varphi_0 \cdot \cos \lambda)$$
 és

$$\lambda = \arcsin \frac{\cos \varphi' \cdot \sin \lambda'}{\cos \varphi}$$

képletekkel számoljuk.

A Gauss- Krüger vetület adatait a térképtanban tanultak alapján számoljuk.

Az 1940-es évek elején a katonai hadvezetés a Gauss-Krüger vetület bevezetése mellett döntött. Az elhelyezést olyan elsőrendű pontok felhasználásával végezték, amelyeknek a közép-európai elhelyezésű és a hazai elhelyezésű Bessel-ellipszoidi koordinátái ismertek voltak. A Gellért-hegy pontra az alábbi értékeket fogadták el:

$$\varphi = 47^\circ 29' 15,382''$$

$$\lambda = 19^\circ 02' 59,723'' \quad (\text{Greenwichtől})$$

$$\lambda = 36^\circ 42' 45,743'' \quad (\text{Ferrotól})$$

A Varsói Szerződés vetületi rendszereként a Gauss-Krüger vetületet fogadták el *Kraszovszkij-ellipszoid* alapfelülettel:

$$a = 6\,378\,245 \text{ m,}$$

$$b = 6\,356\,863,018\,77 \text{ m-es féltengelyekkel.}$$

Az  $x$  tengely irányában nem alkalmaztak eltolást ( $X = x$ ), a negatív előjelű koordináták kiküszöbölése céljából az  $Y$  számításánál mindkét sávban

$$Y_0 = 500\,000 \text{ m}$$

az eltolás mértéke. A sávok megkülönböztetése miatt a 33. sávban **3**-as, a 34. sávban **4**-es *vezérszámot* írnak a  $Y$  koordináta elé.

1969–1975 között új geodéziai vonatkozási rendszer került bevezetésre. Az *IUGG1967 ellipszoid* a Nemzetközi Geodéziai és Geofizikai Unió által 1967-ben elfogadott alapfelület, amelynek jelölésére használatos a *GRS67*. Ez a vonatkozási rendszer később a *HD-72 (Hungarian Datum 1972.)* elnevezést kapta. Az ellipszoidról a síkra a hagyományoknak megfelelően akkor is kettős vetítéssel tértek át. Első lépésben az *IUGG1967 ellipszoid*ról az ellipszoid gömbi vetületével az új Gauss-gömbre, onnan pedig egyetlen ferdetengelyű, két hossztartó segédparalelkörű, (redukált) szögtartó hengervetülettel a síkra. Ez utóbbit nevezük *Egységes országos vetületnek (EOV)* (Vetületi szabályzat 1975). Az *EOV* vetületi kezdőpontja az új Gauss-gömb gellérthegyi meridiánján (kezdőmeridián)

$$\varphi_0 = 47^\circ 06' 0,0000''$$

földrajzi szélességű pontja. A földrajzi hosszúságokat itt is a gellérthegyi meridiántól mérjük (előjelük is keletre pozitív). A kezdőmeridián képe adja az  $x$  tengelyt, a kezdőpontban a kezdőmeridiánra merőleges gömbi főkör (segédegyenlí-

tő) képe az  $y$  tengely. Az  $x$  tengely pozitív ága északra, az  $y$  tengelyé keletre mutat (ÉK-i tájékozású rendszer).

Gyakorlati okokból a síkkoordináta-rendszer kezdőpontját az ország területén kívül eső, alkalmasan kiválasztott pontba helyezték át. Az eredeti  $(y, x)$  vetületi koordináták és a kezdőpont áthelyezése eredményeként nyert  $(Y, X)$  síkkoordináták közötti összefüggés:

$$\begin{aligned} X &= x + X_o, \\ Y &= y + Y_o, \end{aligned}$$

ahol

$$\begin{aligned} X_o &= 200\,000 \text{ m}, \\ Y_o &= 650\,000 \text{ m}. \end{aligned}$$

Így az  $X$  és  $Y$  koordináták mindig pozitív előjelűek, és teljesül az  $X < 400\,000 \text{ m} < Y$

feltétel. Tehát az eltolt koordináták abszolút értékéből egyértelműen megállapítható, hogy  $X$  vagy  $Y$  koordinátáról van-e szó. A távolság- és irányszög számításon kívül természetesen minden vetületi számítás csak az  $x, y$  koordináta párral végezhető.

*IUGG1967* ellipszoidról *EOV* koordináták számításához először az új *ket-tős vetítés* állandóit felhasználva kapjuk az új *Gauss-gömbi* koordinátákat, majd az *EOV* vetületi kezdőpontjától függő segédföldrajzi koordinátákat, végül az *EOV* síkkoordinátákat:

$$\begin{aligned} x &= R \cdot m_o \cdot \ln \operatorname{tg} \left( 45^\circ + \frac{\varphi'}{2} \right) \\ y &= R \cdot m_o \cdot \frac{\lambda'}{\rho^\circ} \quad \rho^\circ = \frac{180^\circ}{\pi} \end{aligned}$$

Az  $X$  és  $Y$  koordinátákat eltolással kapjuk.

Az érintő elhelyezésre vonatkozó koordinátákat az  $m_o = 0,99993$  *vetületi méretarány-tényezővel* szorozzuk, amely megegyezik a segédegyenlítő lineármódulusával és a torzulásmentes segédparalelkörök földrajzi szélességének koszinuszával ( $\cos \phi_{\Xi_m}$ ).

Ha az *EOV* síkkoordinátákból *IUGG1967* ellipszoidi koordinátákat akarunk számítani, akkor első lépésben – mint bármely vetületi számításnál – visszatérünk az eltolás nélküli, vetületi koordinátákra:

$$\begin{aligned} x &= X - X_o, \\ y &= Y - Y_o. \end{aligned}$$

Ezután következik a segédföldrajzi koordináták

$$\varphi' = 2 \cdot \operatorname{arc} \operatorname{tg} e^{\left( \frac{x}{R \cdot m_o} \right)} - 90^\circ = 2 \cdot \operatorname{arc} \operatorname{tg} \exp \left( \frac{x}{R \cdot m_o} \right) - 90^\circ$$



$$\lambda' = \frac{y}{R \cdot m_o} \rho^o$$

majd a gömbi földrajzi koordináták, végül az *IUGG1967* ellipszoidi koordináták számítása.

Az általánosan használt Bessel ellipszoid alapú régi vetületi rendszerek (STG, HÉR, HKR, HDR, BÖV) és az EOV rendszer között közvetlen átszámítási lehetőségre nincs mód. Ezért a transzformáció a Vetületi Szabályzat szerint azonos pontok alapján, ötöd fokú konform hatványsorokkal, a FÖMI-ben történik, az itt kifejlesztett TRAFO nevű transzformációs célprogram segítségével. A TRAFO program a Magyarországon általánosan használt fenti, polgári vetületi rendszerek közötti transzformálásra alkalmas minden kombinációban.

Az *UTM vetület* az ellipszoid egyenlítői elhelyezésű (*transzverzális*) redukált, szögtartó hengervetülete. A vetítés a Gauss–Krüger vetületnél megismert módon történik az ellipszoid felszínéről a hengerpalástra, majd annak kiterítésével a síkra. A torzulásmentes vonalak között a hosszak csökkennek, rajtuk kívül pedig nőnek. A torzulásmentes vonalak a sáv középmeridiánjától  $\Delta\lambda_o = 1^\circ 37' 15''$  -re levő pontokban metszik az egyenlítőt. A síkkoordináta-rendszerek is hasonlóképpen alakulnak, mint a *Gauss–Krüger* vetületnél. Az érintő elhelyezéshez tartozó síkkoordinátákat az

$$m_o = \cos \Delta\lambda_o = 0,9996$$

vetületi méretarány-tényezővel szorozva kapjuk az *UTM* koordinátákat.

Az *UTM* vetületi számításoknál a *Gauss-Krüger* vetület összefüggéseit használhatjuk, néhány jelölésbeli különbséggel (ld. *táblázat*):

Gauss - Krüger vetület	UTM vetület
$m_o = 1$	$m_o = 0,9996$
$x$	$\Delta N$
$y$	$\Delta E$
$X$	$N$
$Y$	$E$
$X_o$	$FN$ (False Northing)
$Y_o$	$FE$ (False Easting)

A Gauss-Krüger vetületnél az  $X_o = 0$ , az UTM *FN* eltolás értéke csak az egyenlítőtől északra 0 (így Magyarországon is), délre pedig 10 000 000 m. A *Gauss-Krüger*  $Y_o$  eltolás értéke mindegyik sávban 500 000 m, ezért a sáv megkülönböztetése érdekében a  $Y$  koordináta elé az ún. *vezérszámot* (a 33. sávnál **3**, a 34. sávnál **4**) írjuk. Az UTM *FE* eltolás értéke is mindegyik sávban 500 000 m), így a sáv meghatározásához itt is szükségünk van még egy adatra, nevezetesen annak a  $6^\circ \times 8^\circ$  -os ellipszoidi négyszögnek a jelzésére, amelyikben a pontunk

elhelyezkedik. Magyarország területén ezeknek a négyszögeknek az elhatároló vonalai ugyanazok, mint a Gauss-Krüger szelvényezésnél, csak másik ellipszoidon. Itt  $33 T$ ,  $33 U$ ,  $34 T$  vagy  $34 U$  jelzéseket írjuk a pont helyzetétől függően az  $E$  koordináta elé vagy mögé.

A GPS technika alkalmazása új egységes koordináta-rendszert kíván meg, amit Magyarországon a katonai topográfiában a WGS84 ellipszoidhoz mint alapfelülethez tartozó UTM vetület biztosít. A nyugat-európai országokban általában a Hayford-féle ellipszoidot alkalmazzák alapfelületként. A sáv szélesség  $6^\circ$ -os és a középmeridiánok ugyanazok, mint a Gauss-Krüger vetületnél.

Az UTM vetületi számításokhoz is a Gauss-Krüger vetületnél megismert összefüggéseket kell használni, azzal a különbséggel, hogy a Gauss-Krüger vetületnél az  $m_0$  helyébe 1-et, az UTM-nél 0,9996-ot kell helyettesíteni, valamint a számítható állandókat az elfogadott alapfelülethez kell meghatározni. Az előbbi állandóknak a gyakorlati követelményeknek megfelelő pontossággal kiszámított értékei a WGS84 ellipszoidra:

$$R = 6\,367\,449,149\text{m};$$

$$\alpha_2 = 8,377\,318\text{ EE-04}; \alpha_4 = 7,608\,5\text{ EE-07}; \alpha_6 = 1,2\text{ EE-09}; \alpha_8 = 0;$$

$$\beta_2 = 8,377\,322\text{ EE-04}; \beta_4 = 5,9\text{ EE-08}; \beta_6 = 2\text{ EE-10}; \beta_8 = 0$$

### ***Feladatok***

- 1, Az első, második és harmadik katonai felvételeken határozzuk meg Debrecen középpontjának koordinátáit!
- 2, Határozzuk meg a Budai rendszer kezdőpontjának EOVS koordinátáit!
- 3, Határozzuk meg egy tetszőleges hazai pont EOVS koordinátáiból az IUGG1967 ellipszoidi koordinátáit!

## 5. Adatok, adattípusok, adatgyűjtés (minta, mintavételezés és szabályai). Régi mértékegységek és átszámításuk

A mindennapi életben különféle adatokkal találkozunk (pl. személyi, statisztikai, közérdekű, műszaki stb.). Ebben a fejezetben először a geoinformatikában gyakran előforduló adatok tipizálásával foglalkozunk, majd megismerjük a különböző adattípusok gyűjtésének lehetőségeit. Áttekintjük a mintaválasztás és a mintavételezés szabályait. Mivel az adatokhoz valamilyen mértékegységek is kapcsolódnak, és ezek a történelem során többször változtak, ezért átismétljük a régi és az új mértékegységekkel kapcsolatos tudnivalókat, az átváltásuk lehetőségeit, korlátait.

Az adatok feldolgozásával tulajdonképpen az a célunk, hogy újabb információkhoz jussunk. A feldolgozás lehet nagyon egyszerű (pl. alapl műveletek, átlag, stb.), amelyek számológépekkel is elvégezhetők. Ennél összetettebb műveleteket végezhetünk táblázatkezelő programokkal (pl. EXCEL), illetve a bonyolultabb feladatok megoldásához különféle tudományos és GIS programokat használhatunk.

Az adatfeldolgozás egy sajátos területe az adatbáziskezelés, amelyet adatbáziskezelő szoftverekkel végezhetünk. Ezzel a témakörrel külön tantárgy foglalkozik.

### 5. 1. Adatok, adattípusok

Az adatoknak (pl. 80, zöld) önmagukban nincs jelentésük. Az adatoknak csak akkor van értelme, ha egy objektum tulajdonságának, jellemzőjének (változójának) az értékét állapotát jelzik (pl. 80 Ft, zöld alma, stb.). Tehát egy adatot akkor tekintünk definiáltnak, ha meghatározzuk, hogy milyen objektum, melyik változója, milyen értéket vesz fel. Így már hasznos adattá, információhordozóvá válik. Az adatokat különféle szempontok szerint csoportosíthatjuk.

Az adatokat tipizálhatjuk az elvégezhető műveletek szerint:

- a **nominális adat** lehetséges értékei között csak az azonos vagy nem azonos reláció van értelmezve (egyenlő vagy nem egyenlő). Ilyenek például a nevek (Debrecen, Budapest, Petőfi Sándor), telefonszámok (36-52-512900, 36-52-480555), színekkel jellemzett tulajdonságok (zöld alma, piros kör). Nominális adattal ennél magasabb rendű műveletet nem lehet végezni, ebből következően (Debrecenből nem vonható ki Szeged és a telefonszámok összege sem értelmezhető). Az adatbázisban leggyakrabban szöveges változóként jelennek meg. Éppen ezért, bizonyos műveletek esetén (pl. csoportosítás, lásd diszkriminancia-analízis, SPSS), át kell őket

definiálni szám típusú változóvá (Debrecen = 1, Szeged = 2, de ezek a számok csupán a csoportok elkülönítésére szolgálnak, nem ordinális és nem metrikus arányskálán mért értékek). Ez a módosítás szükségszerű, de végeredményben önkényes.

- az **ordinális adat** lehetséges értékei között a kisebb vagy nagyobb reláció is megengedett az azonos nem azonos mellett (=, ≠, <, >). Ilyen pl. < 1000 mg, 1000 – 1500 mg. Az osztályozás során ordinális adatok keletkeznek.
- az **intervallum skálán mért adatok**nál a fentiekén kívül az összeadás és kivonás műveletét is tudjuk értelmezni, vagyis megadható, hogy mennyivel kisebb vagy nagyobb két érték egymástól, de a szorzás és osztás művelete nincs értelmezve. Ilyen például a Celsius fokokban mért hőmérséklet, mert skálájának 0 pontja önkényes. (A 40°C-ról mondhatjuk, hogy 20°C-kal melegebb a 20°C-nál, de azt nem mondhatjuk, hogy kétszer melegebb.)
- az **arányskálán mért adatok**nál mind a négy alpművelet értelmezve van. Például a kilogrammban mért tömeg, newtonban mért erő, darabban vagy mólbán mért elemszám, kelvinben mért hőmérséklet, méterben mért távolság, stb.

Az adatbáziskezelőkben előforduló adattípusok:

- **szöveg (Text) típusú adatok:** Akkor használjuk, ha a beírt adattal nem kell számításokat végezni (pl. nevek, irányítószámok, stb.)
- **feljegyzés (Memo) típusú adatok:** Akkor használjuk, ha a mezőben 255 karakternél hosszabb szöveget szeretnénk tárolni.
- **számtípusú (Number) adatok:** Matematikai számítások elvégzésére alkalmas adatok. Egész és törtszámok egyaránt lehetnek.
- **pénznem típusú adatok:** Pénzértékek, amelyek 15 egész és négy tizedes jegyet tartalmazhatnak. Különböző pénznemek adhatók meg.
- **dátum/idő (Date/Time) típusú adatok:** Akkor használjuk, ha dátumokkal és időpontokkal akarunk műveleteket végezni (pl. ki kell számolni két dátum között eltelt napok számát). A napok végére nem szabad pontot írni.
- **sorszám (Auto Number) típusú adatok:** Egyértelmű egész számértékek, amelyeknél beállítható a növekvő/csökkenő sorrend. Adattáblák összekapcsolásánál játszhat szerepet, mint elsődleges azonosító (kapcsolómező), de használatával körültekintően kell eljárni. Helyette saját készítésű azonosító bevezetése javasolt.
- **igen/nem (Yes/No, True/False) típusú adatok:** Ezek tulajdonképpen logikai értékek (igen/nem, igaz/hamis). Akkor használjuk, ha előre eldönthető, hogy az adott mező csak kétféle értéket vehet fel.

- **OLE objektum adattípusok:** Akkor alkalmazzuk, ha egy másik programmal előállított objektumokat akarunk az adatbázishoz csatolni. Ilyenek pl. a kép, zene, film és más bináris adatok. Lehetőség van hyperlink ([www.unideb.hu](http://www.unideb.hu)) megadására is.

A geoinformatikai szoftvereknél gyakori a **Number, String, Boolean és Date** típusú adatok használata.

- A **Number** (numerikus – szám) adattípus választásánál előre meg kell határoznunk, hogy az adatmező maximálisan hány értékes számjegyet tartalmazhat, és a tizedesjegyek számát is. Ennek alapja egyes porogramoknál gyakran a kettes számrendszer.
- A **String** választásakor a szöveges információk tárolására van lehetőség. Ebben az esetben a karakterek számát kell előre meghatározni.
- A **Boolean** változó szintén kétértékű (I/H), így az adattároláson felül bizonyos logikai műveletek elvégzésére is alkalmas.
- A **Date** típusú adatok a dátummal, idővel kapcsolatos feladatok megoldását támogatja.

## 5. 2. Adatgyűjtés, adatnyerési eljárások, adatforrások

Az objektumok helyzete, attribútuma időben (és térben) folyamatosan változik, ezért az adatbázisunkat állandóan frissíteni kell. (pl. a lakóházak mérete hozzáépítéssel megváltozik, vagy tulajdonosváltás történt, stb.) A változás elemzésénél megállapíthatjuk, hogy a geometriai objektum alakja hogyan változott, és milyen új attribútum adatokat kell hozzárendelni. A változás időpontját is fel kell tüntetni. Fontos az adat pontosságának (minőségének) az ellenőrzése. A nem megfelelő adatminőség rossz eredményt és ennek következtében hibás döntést hozhat. A hibás döntés a gyakorlati életben jogi és pénzügyi következményekkel járhat.

A kutatás során az adatok gyűjtésénél és a mintavételezésnél nagyon sokféle módszert alkalmaznak, ezért a szabályzatok, előírások is nagyon eltérőek. Itt terjedelmi okok miatt nem célunk az összes módszer tárgyalása, sőt a geográfia-ban előfordulók közül is csak néhány bemutatásával érzékeltetjük az adatgyűjtés lehetőségeinek a sokszínűségét.

A geográfia-nak bármely tudományterületén nélkülözhetetlen a térkép. A tematikus térképek előállításához nagy mennyiségű adatra van szükség. A napjainkban készülő digitális tematikus térképek a terepi objektumok adatállományát numerikus kódolással tartalmazzák.

A digitális térképek adatainak előállítása során terepi felmérést, fotogrammetriai eljárást, vagy vegyes eljárást alkalmazunk. A terepi felmérésnél az objektumok alakjelző pontjainak helyszíni bemérését végzik az elektronikus

tahiméterekkel, vagy mérőállomásokkal. A helymeghatározás olyan GPS-el készül, amely ma már részletes felmérésre is alkalmas.

Újabban a „kézi számítógépek” használatának elterjedésével az ArcPad szoftver használata lehetőséget nyújt a terepen való elemfelvételre, elemzésre és a földrajzi információk megjelenítésére. A terepi adatgyűjtés az ArcPad-dal hatékony és megbízható terepi adatgyűjtés. Lehetőségünk van külső szenzorokból érkező adatok (GPS vevők, távmérők és digitális kamerák) felvételére is.

A műszerek a mérési eredményeket digitális formában rögzítik és az adatok a további feldolgozás céljából közvetlenül számítógépbe tölthetők. A feldolgozást megfelelő szoftverrel elvégezve az adatokból a digitális térkép megszerkeszthető.

A térképi objektumok attribútum adatai különféle adatbázisokból (pl. KSH) letölthetők, illetve korábban készült táblázatokból, felmérésekből, stb. összegyűjthetők, vagy megvásárolhatók. Napjainkban az adatgyűjtést jelentősen gyorsítja, és könnyebbé teszi az INTERNET.

A digitális térképek előállítását gyakran a régi térképek adatállományának frissítésével végezzük. Ekkor jól alkalmazhatók a légi- és úrfelvételek. Ennél a módszernél nagyon fontos az új adatbázis terepi ellenőrzése.

Abban az esetben, ha rendelkezésünkre áll megfelelő digitális térkép és az objektumok adatai, akkor a különféle tematikus térképek geoinformatikai szoftverekkel gyorsan előállíthatók.

### 5. 3. Mintavételezés

Az új adatok gyűjtése, előállítása történhet mintavételezéssel. A mintavételezésnél a különböző szakterületek eltérő módszereket és szabályokat alkalmaznak. A mintavételezés lehet terepi (pl. talajminták gyűjtése, kérdőívezés stb.), vagy adatbázisból történő adatválogatással. A földtudományokban a mintavételezésnek ezeket a típusait gyakran alkalmazzuk. A mintavételezés **reprezentatív**, ha a minta a statisztikai sokaság jellemzőit viseli magán (pl. a mintában a nyugdíjasok aránya megegyezik a sokaságban a nyugdíjasok arányával; vagy a minták, mint részhalmaz átlaga hasonló az eredeti halmaz, a sokaság átlagához). A mintavételezés nem szükségszerűen törekszik reprezentativitásra: pl. elemezhető a diákok pártpreferenciája és külön a nyugdíjasoké, s e két részhalmaz vizsgálata vélhetőleg más eredményt ad, mintha az összes választópolgár pártpreferenciáit vizsgálnánk. A sokaság nagysága befolyásolja a reprezentativitáshoz szükséges mintaszámot: 10000 eset felett elegendő, ha a minta az esetek 1%-át tartalmazza, ennél kisebb nagyságrend esetében a minta arányát/számát növelni kell.

A geográfiaiban a terepi mintavétel, mint adatnyerési eljárás sokszor nélkülözhetetlen. A mintavétel helyének kiválasztásához magas szintű szakmai fel-

készültség szükséges. A mintavételezés módjai nagyon eltérőek és az eredmény kimenetelét is befolyásolják.

A mintavétel helyének kiválasztása az alábbi módon történhet:

- Találomra – szubjektív, nem ajánlott.
- Véletlenszerű kijelölés – nem biztos, hogy reprezentatív.
- Szisztematikus kijelölés – általában négyzethálós, sok fölösleges pont.
- Szisztematikus – véletlen – a négyzetrácson belül véletlenszerű kijelöléssel.
- Digitális terepi modellek, légi- és űrfelvételek segítségével.
- Homogén foltok meghatározása – becslésre jó, de mintázatelemzésre alkalmatlan.
- Tér folyamat-függő hálózat optimalizálásával.

A helyek kiválasztásánál szokták még alkalmazni a krigelés becslési varianciáján alapuló módszert, amelynél az exponenciális kovariancia függvényt (covariogram) használják. Egy adott területen a lehetséges hálózatok közül az a hálózat a legjobb, amelynél ez a kritérium a legkisebb értékkel rendelkezik.

A kovariancia függvény alkalmazásánál felvetődik néhány kérdés pl.:

- Hogyan értelmezhetőek a kovariancia függvény paraméterei a gyakorlat számára?
- Milyen hatása van a paramétereknek a kritérium értékére?
- Megadható-e a hálózathoz kiválasztott helyek optimális száma?
- Változnak-e a kiválasztott helyek, ha a vizsgált területen belül valamilyen egyéb hatást, heterogenitást is feltételezünk?

A nagy területek mintavételezése a térbeli heterogenitás elemzése céljából mindig problémát jelent. A feladatok figyelembe vételével általában mindig új mintavételezési eljárások kidolgozására van szükség. Ezeknél az eljárásoknál a terepi mintavételezés mellett célszerű új modern módszereket (távérzékelés, modellezés) is alkalmazni. *A mintavétel után szükséges a hibaelemzés, a pontosság-vizsgálat, és az adatok minőségének az értékelése.*

#### 5. 4. Műveletek adatokkal

Az adatgyűjtés/mintavételezés után az információszerzés érdekében a rendelkezésünkre álló adathalmazzal különféle műveleteket végzünk. Az adatokat először rendezzük, osztályozzuk.

Az adatainkat **nominális változóknak** tekintjük akkor, ha csak azt tudjuk megadni, hogy az egyes osztályokba (kategóriákba) hány egyed esett. Jelölje pl.  $n_i$  azoknak az egyedeknek a számát, amelyek az  $i$ -edik kategóriához tartoznak.

Ekkor az  $n_i$ -t *gyakoriságnak* nevezzük. A relatív gyakoriság ( $f_i = n_i/N$ ) pedig azt mutatja meg, hogy az adataink hány százaléka esett az  $i$ -edik kategóriába (pl.  $f_i=0,65$  értéknél 65%).

Adataink gyakran egy, vagy több *intervallumba* eső értékek. Ilyenkor célszerű az adatokat nagyság szerint sorba rendezni, a kategória határokat kijelölni és a különböző osztályokba besorolni ( $n_i$  és  $f_i$  értékeket meghatározni). Adataink megoszlásának szemléltetésére *hisztogramot* (gyakorisági, eloszlásgörbét) lehet szerkeszteni derékszögű koordináta rendszerben. (Az egyik tengelyre az osztályközépeket, a másikra pedig a hozzájuk tartozó gyakoriságot, illetve relatív gyakoriságot mérjük fel.) A hisztogram tájékoztat bennünket a változó eloszlásáról, csúcsairól, szimmetriájáról, stb.

Az adatok értékelésénél fontos a középérték és a szórás mérőszámainak az ismerete. Ezek a *statisztikai paraméterek* ugyanis bizonyos eloszlások esetén reprezentálhatják az egész sokaságot (vagy részhalmazát, a mintát), így összevetésükkel több adathalmaz, populáció összehasonlítására is lehetőség nyílik, ami jóval könnyebb, mint az összes, akár több ezer adatot összevetni egymással. Használatuk átvezet a statisztika és a valószínűségszámítás és a halmazműveletek területeire (lásd ott).

#### A középérték mérőszámai:

- A **modus** ( $M_0$ ) arra ad választ, hogy melyik a változó legvalószínűbb értéke. Tulajdonképpen a hisztogram maximumát jelenti. Ennek alapján lehetnek többmodusú, illetve modulus nélküli változók is.
- A **medián** ( $M_e$  – középső érték) pontosan felezi a mintát, tehát az itt húzott vonal az eloszlásgörbe alatti terület felezője.
- A **számtani átlag** ( $\bar{x}$ ) a változó gyorsan meghatározható, egyetlen számmal történő jellemzése. Képlete:

$$\bar{x} = \frac{\sum_{i=1}^k x_i}{N} = \frac{\sum_{i=1}^k n_i \cdot x_i}{\sum_{i=1}^k n_i}$$

ahol  $k$  a csoportok száma,  $n_i$  az  $i$ -edik csoport gyakorisága

$$\sum_{i=1}^k n_i = N$$

$n$  a minta nagysága

(Lehet súlyozott átlagot is számolni, de akkor az  $n_i$ -nek más az értelmezése.)



- **Harmonikus átlag** ( $x_h$ )

$$\bar{x}_h = \frac{N}{\sum_{i=1}^N \frac{1}{x_i}}$$

- **Mértani (geometriai) átlag** ( $x_g$ )

$$\bar{x}_g = \sqrt[N]{x_1 \cdot x_2 \cdot \dots \cdot x_N}$$

- **Négyzetes átlag** ( $x_q$ )

$$\bar{x}_q = \sqrt{\frac{\sum_{i=1}^N x_i^2}{N}}$$

A különböző átlagok csak a változó (eloszlás) egy jellemző értékét adják meg. Az intervallumban található változóknál célszerű meghatározni az értékek szóródását is.

**A változók szóródásának mérőszámai:**

- **Terjedelem** ( $R$ ):  $R = X_{\max} - X_{\min}$

- **Interkvartilis féltérjedelem** ( $IF$ )

$$IF = \frac{K_3 - K_1}{2}$$

ahol  $K_3$  a harmadik,  $K_1$  az első kvartilis (negyed)

- **Átlagos eltérés** ( $\delta$ ) – a számtani átlagtól való eltérések abszolút értékének átlaga

$$\delta = \frac{\sum_{i=1}^N |x_i - \bar{x}|}{N}$$

- **Szórás (variancia, kovariancia)** – a számtani átlagtól való eltérések négyzetes átlaga.

a szórásnégyzet képlete:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}} \quad \sigma^2 = \frac{1}{N} \cdot \sum_{i=1}^N (x_i - \bar{x})^2$$

A regionális elemzésekben használatosak statisztikai paramétereken alapuló területi egyenlőtlenségi mutatók (vö: Regionális elemzési módszerek; Péntes J. előadásai alapján).

## Területi megoszlások eltérését mutató indexek:

a; Koncentrációs-index (Hirschman-Herfindahl-index)

Képlete	Jelölések	Értelmezése
$K = \sum_{i=1}^n \left( \frac{x_i}{\sum_{i=1}^n x_i} \right)^2$	n = az adatsor elemszáma	Valamely természetes jellemző terület-egységek közötti koncentrációjának mértékét számszerűsíti. A megoszlást az index a teljesen egyenleteshez (amikor minden megfigyelési egység részesedése azonos) viszonyítja. 0,6 feletti értéke már erős koncentrátságra, monopólihelyzetre utal.

Mértékegysége: dimenziótlan Értékkészlete:  $1/n \leq K \leq 1$

**Megjegyzések:** Elterjedt mutatószám a területi kutatásokban, minimális értékét teljesen homogén eloszlás, maximális értékét pedig teljes (egy kézben történő) koncentráció esetén veszi fel. Mivel a minimum értéke erősen függ az elemszámtól, eltérő elemszámú vizsgálatok esetén az eredményeket nem lehet közvetlenül összehasonlítani. Ezért minimumértéke nem 0.

c; Redundancia mutató vagy Theil-index

Képlete	Jelölések	Értelmezése
$R = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{\bar{y}} \log \left( \frac{y_i}{\bar{y}} \right)$	$y_i = \frac{x_i}{f_i}$ fajlagos mutató értéke az i. területegységben $\bar{y} = y_i$ súlyozott átlaga	A Theil-index az entrópia koncepciójára épül és a vizsgált ismérv össz-volumenéből való részesedések rendezetlenségét méri.

Mértékegysége: dimenziótlan Értékkészlete:  $0 \leq R \leq \log n$

**Megjegyzések:** minimális értékét akkor veszi fel, ha minden fajlagos (pl. a jövedelem) érték azonos, maximumát pedig akkor, ha a vizsgált ismérv egy „kézben”, egy területegységben összpontosul. A logaritmus alapja szerint különböző indexeket lehet számítani. Leggyakrabban a 10-es és a természetes alapú logaritmusokat használják.

e; Gini-együttható (Gini-index)

Képlete	Jelölések	Értelmezése
$G = \frac{1}{2 \cdot n^2 \cdot \bar{y}^2} \sum_{i,j}  y_i - y_j $	$y_i / y_j =$ megoszlási viszonyozóminté megadott területi jellemző az i./j. területegységben $\bar{y} = x_i$ átlaga	A Lorenz-görbe és az átló által bezárt terület nagyságát méri, a koncentráció relatív nagyságát jellemzi. Minden megfigyelési egység részarányának az összes többiéitől való átlagos eltérését viszonyítja az átlaghoz.

Mértékegysége: dimenziótlan Értékkészlete:  $0 \leq G \leq 1$

**Megjegyzések:** A 0 értéket akkor veszi fel, ha a Lorenz-görbe éppen egybeesik az átlóval, tehát a vizsgált mennyiségi ismérv területi eloszlása egyenletes. Másik szélső értéket akkor éri el, ha a vizsgált ismérv egyetlen „kézben” összpontosul; ilyenkor a görbe egybeesik a koordinátengelyekkel. A jövedelemegyenlőtlenségek mérésének legelterjedtebb mutatója.

## A területi polarizáltság mérőszámai

b; Duál-mutató (Éltető-Frigyes index)

Képlete	Jelölések	Értelmezése
$D = \frac{x_m}{x_a}$	$x_m =$ az $\bar{x}$ értékénél nagyobb $x_i$ értékek számtani/súlyozott számtani átlaga $x_a =$ az $\bar{x}$ értékénél kisebb $x_i$ értékek számtani/súlyozott számtani átlaga	A duál-mutató a teljes megoszlás átlaga fölötti értékek átlagának és a teljes megoszlás átlaga alatti értékek átlagának hányadosa. Egyszerűsége és világos tartalma miatt igen elterjedt módszer.

Mértékegysége: dimenziótlan Értékkészlete:  $1 \leq D < \infty$

**Megjegyzések:** Pl. jövedelemegyenlőség esetén az értéke 1, ennél nagyobb érték esetén az index azt a jövedelmi ollót mutatja, amely az átlagosan gazdagok (átlag felettek) és az átlagosan szegényebbek (átlag alattiak) jövedelme között fennáll. A mutató jó szolgálatot tesz a területi egyenlőtlenségek tényezőkre bontásakor is.

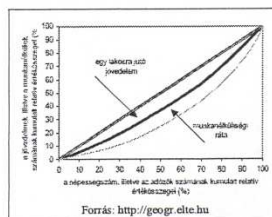
b; Hoover-index (Robin Hood-index)

Képlete	Jelölések	Értelmezése
$h = \frac{\sum_{i=1}^n  x_i - f_i }{2}$	$x_i$ és $f_i$ két megoszlási viszonyozóminté fennállnak az alábbi összefüggések: $\sum x_i = 100$ $\sum f_i = 100$	A Hoover-index két mennyiségi ismérv területi megoszlásának eltérését méri. A mutató szimmetrikus, a két összevetett megoszlás szerepe, sorrendje felcserélhető. (Krugman-index)

Mértékegysége: % Értékkészlete:  $0 \leq h \leq 100$

**Megjegyzések:** A Hoover-index az egyik legelterjedtebb, általánosan használt területi egyenlőtlenségi mutató. Azt adja meg, hogy az egyik gazdasági-társadalmi jelenség mennyiségének hány százalékát kell átcsoportosítani ahhoz, hogy területi megoszlása a másik jellemzőével azonos legyen. A területi kutatásokban leggyakrabban a népesség területi eloszlásával vetik össze különböző társadalmi-gazdasági tartalommal bíró mennyiségi ismérvek eloszlását.

d; Lorenz-görbe



A kumulált relatív gyakoriságok függvényében ábrázolja a kumulált relatív értékösszegeket. Ha az egységeknek értékösszege azonos, akkor a kumulált relatív gyakoriságok és a kumulált értékösszegek rendre megegyeznek, és ebben az esetben a koncentráció hiánya áll fenn, tehát a görbe megegyezik a négyzet átlójával.

A Lorenz-görbe előnye a területi egyenlőtlenségek vizuális megjelenítése, alkalmazásának gátja azonban, hogy a görbék metszhetik egymást – vagyis a két eloszlás jellege eltérő lehet.

f; Súlyozott Gini-együttható

Képlete	Értelmezése
$G_s = \frac{1}{2 \bar{y}_s} \sum_{i,j} \frac{f_i f_j}{(\sum f)^2}  y_i - y_j $	A Gini-koefficiens súlyozott változata is a Lorenz-görbe által bezárt területtel arányos, viszont olyan görbét kell elképzelnünk, ahol a vizsgált fajlagos mutató két összevetője közül az egyik kumulált relatív gyakoriságainak függvényében ábrázolja a másik kumulált relatív értékösszegeit.

Mértékegysége: dimenziótlan Értékkészlete:  $0 \leq G_s \leq 1$

**Megjegyzések:** Gyakran használatos, hiszen ezáltal fajlagos mutatókat, pl. egy főre eső mutatókat lehet számolni.

b; Relatív szórás

Képlete	Jelölések	Értelmezése
$V = 100 \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n \bar{x}^2}}$	$n =$ az adatsor elemszáma $\bar{x} = x_i$ számtani átlaga	A relatív szórás a szórás vizsgált adatsorának átlagához viszonyított mértékét jelzi.

Mértékegysége: % Értékkészlete:  $0 \leq V < \infty$

**Megjegyzések:** Sok esetben szükség lehet arra, hogy összehasonlíthatóvá tegyük a különböző mértékegységű jelzőszámok szórását. Ekor szokás valamelyik középértékhez viszonyítani. Más néven variációs koefficiensnek vagy standard deviációnak nevezik. 100-zal szorozva %-ban fejezi ki a szórás.

**Súlyozott szórás:**

$$\sigma_s = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2 f_i}{\sum_{i=1}^n f_i}}$$

$y_i = \frac{x_i}{f_i}$  fajlagos (arány) mutató értéke az i. területegységben  
 $\bar{y} = y_i$  súlyozott átlaga  
 $f_i$  = súlyszám (a területi társadalmi-gazdasági vizsgálatokban általában a lakosság szám, de jellemzően a fajlagos mutatóknak megfelelően változhat)

A közgazdaságban  $\sigma$ -konvergenciának nevezik a szórás alapján csökkenő jövedelem-egyenlőtlenségeket.. Csak azonos mértékegységű jellemzők vehetők össze.

**Példa:**

	Foglalkoztatottsági ráta, %, 1990	Lakónépesség száma, fő, 1990	Foglalkoztatottsági ráta, %, 2001	Lakónépesség száma, fő, 2001
Bács-Kiskun	43,7	546898	35,0	546517
Baranya	42,8	417400	33,5	407448
Békés	42,0	411887	31,5	397791
Borsod-Abaúj-Zemplén	41,6	761963	28,1	744404
Budapest	45,5	2016681	42,0	1776388
Csongrád	43,4	438842	35,6	433344
Fejér	45,8	418852	40,4	434317
Győr-Moson-Sopron	44,7	428153	42,1	434706
Hajdú-Bihar	41,4	548728	31,6	552998
Heves	42,6	334408	33,8	325727
Jász-Nagykun-Szolnok	42,6	425583	32,1	415917
Komárom-Esztergom	44,2	316984	39,2	316590
Nógrád	43,1	227137	33,0	220261
Pest	45,1	949842	38,4	1083877
Somogy	42,7	344708	33,6	335237
Szabolcs-Szatmár-Bereg	38,5	572301	27,0	582256
Tolna	44,0	253675	35,2	249683
Vas	45,2	275944	42,8	268123
Veszprém	45,0	378439	39,4	373794
Zala	44,3	306398	39,5	297404
Országos	43,6	10374823	36,2	10196782

Forrás: népszámlálási adatok alapján

**Lépések:**

$$\sigma_s = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2 f_i}{\sum_{i=1}^n f_i}} = \sqrt{\frac{[(43,7 - 43,6)^2 \cdot 546898] + [(42,8 - 43,6)^2 \cdot 417400] + \dots + [(44,3 - 43,6)^2 \cdot 306398]}{10374823}} = \sqrt{\frac{36205243}{10374823}} = \sqrt{3,4897} = \underline{\underline{1,868}}$$

A számítás első lépéseként a fajlagos mutató ( $y_i$ ) értékéből kivonjuk az átlag értékét ( $\bar{y}$ ), mely műveletet az adatsor minden elemére elvégezzük ( $y_1$ -től  $y_n$ -

ig). A példában az egyes magyarországi megyék foglalkoztatottsági arány értékéből vonjuk ki az országos átlagértéket. Ezt követően minden különbséget négyzetre emelünk, majd besorozzuk a hozzá tartozó súlyértékkel ( $f_i$ ), mely esetünkben a lakónépesség számát jelenti. A szorzatok összegét képezzük, melyet a súlyértékek összegével osztunk el, s a hányados négyzetgyökét véve megkapjuk a súlyozott szórás értékét.

Amennyiben a kapott értéket elosztjuk az  $\bar{y}$  értékével és besorozzuk 100-zal; a súlyozott relatív szórás értékét kapjuk meg, amely a súlyozott átlag százalékában fejezi ki a szórás értékét. (A súlyozott szórás számítást elvégezve a 2001-es adatokra is, 4,9021-et kapunk végeredményként, mely alapján megállapítható, hogy a foglalkoztatottsági ráta megyék közötti különbségei jelentősen növekedtek a két népszámlálás közötti időszakban).

**Hoover index:**

$$h = \frac{\sum_{i=1}^n |x_i - f_i|}{2}$$

$x_i$  és  $f_i$  két megoszlási viszonyszám, melyekre fennállnak az alábbi összefüggések:  
 $\sum x_i = 100$  és  $\sum f_i = 100$

**Példa:**

Régiók	Bruttó hazai termék (GDP), Mrd Ft-ban, 2000	Lakónépesség száma, fő, 2000	Bruttó hazai termék (GDP), Mrd Ft-ban, 2006	Lakónépesség száma, fő, 2006
Dél-Alföld	1 340	1391786	2 103	1359617
Dél-Dunántúl	994	1005610	1 547	983247
Észak-Alföld	1 340	1586824	2 278	1559233
Észak-Magyarország	1 112	1318098	1 899	1278550
Közép-Dunántúl	1 443	1127597	2 370	1124315
Közép-Magyarország	5 781	2819694	11 229	2858844
Nyugat-Dunántúl	1 519	1004978	2 369	998241
Ország összesen	13 529	10 254 587	23 795	10 162 047

Forrás: a KSH Stadat adatai alapján

A számítás első lépéseként megoszlási viszonyzámmá kell átalakítani a természetes mértékegységben megadott területi adatokat (mind a vizsgált változót, mind pedig az összehasonlítás alapját képező sokaságot – társadalmi-gazdasági vizsgálatok esetében általában a népességszámot), melyekre igaz, hogy  $\sum x_i = 100$  és  $\sum f_i = 100$ . A következő lépésben a két megoszlás különbségét vesszük minden terület egység esetén, majd a különbségek abszolút értékeit összegezzük. A kapott összeget kettővel elosztva megkapjuk a Hoover-index értékét. A 2006-os adatokra is elvégezve a számítást, eredményül 19,19%-ot kapunk, mely alap-

ján megállapítható, hogy a vizsgált időszakban a GDP tekintetében növekedett a régiók közötti egyenlőtlenség.

*Lépések:*

Régiók	Bruttó hazai termék (GDP), Mrd Ft-ban, 2000	Lakónépesség száma, fő, 2000	Bruttó hazai termék (GDP) megoszlása, %	Lakónépesség számának megoszlása, %
Dél-Alföld	1 340	1391786	9,91	13,57
Dél-Dunántúl	994	1005610	7,35	9,81
Észak-Alföld	1 340	1586824	9,91	15,47
Észak-Magyarország	1 112	1318098	8,22	12,85
Közép-Dunántúl	1 443	1127597	10,66	11,00
Közép-Magyarország	5 781	2819694	42,73	27,50
Nyugat-Dunántúl	1 519	1004978	11,23	9,80
Ország összesen	13 529	10 254 587	100,00	100,00

$$h = \frac{\sum_{i=1}^n |x_i - f_i|}{2} = \frac{|9,91 - 13,57| + |7,35 - 9,81| + \dots + |11,23 - 9,80|}{2} = \frac{3,66 + 2,46 + \dots + 1,43}{2} = \frac{33,32}{2} = \underline{\underline{16,66}}$$

Gyakran előfordul, hogy adatsorunkat újra kell skálázni, vagy osztályközöket kell létrehozni. Az intervallum-képzésnek számos módszere ismert, a vizsgálati cél mindig befolyásolja a módszertant. Ilyen az

**a**, egyenlő osztályközök elve – ebben az esetben nem vagyunk tekintettel arra, hogy egy intervallumba hány adat tartozik (pl. az adatterjedelem negyedelése)

**b**, egyenlő adatmennyiség elve – ebben az esetben egyenlő számú adat kerül minden kategóriába (az adatmennyiség negyedelése)

**c**, hisztogram-alapú osztás – ebben az esetben a móduszok önálló kategóriát képviselnek, míg a köztük lévő átmenetek együtt kerülnek osztályba sorolásra.

**d**, az intervallum-határ ott is meghúzható, ahol hiátust vagy nagyobb ugrást látunk, az adatokat növekvő sorrendbe rendezve – ebben az esetben az intervallumok számát a hiátusok száma/nagysága szabja meg, az intervallumok nem azonos nagyságúak és nem azonos mennyiségű adat kerül besorolásra. Alkalmazható szennyezés-érzékenységi térképeknél, ahol az intervallumok minőségi ugrást jelenítenek meg.

**e**, alapvető szabály, hogy legalább 3, maximum 7-8 kategóriát különítsünk el az adatok térképi megjelenítésekor, ugyanis ennél több kategória átláthatatlanná teszi az ábrázolást. Ez az eljárás igaz nemcsak egyszerű változók esetén, de a klaszterek vagy a faktorok számának beállításakor is célszerű figyelembe venni (lásd ott).

**f**, a kategóriák számának megállapításakor figyelembe vehetjük a következő összefüggést is:

$$k = 1 + 3,3 \cdot \lg n \quad (\text{ahol } k \text{ a kategóriák maximális száma, } n \text{ az adatok száma})$$

Előfordulhat, hogy több adatsor értékeit terjedelmi okokból egy, összevont kimeneti térképen kívánjuk megjeleníteni. Ilyenek a különböző pontszámítási rendszerek és az ezeken alapuló térképek: pl. a szennyezés-érzékenységnél figyelembe vehető a szemcseösszetétel, talajvízállás és az agyagásaványtartalom, mint befolyásoló tényezők. Mivel értékük nagyságrendileg eltérő és más-más dimenzióval, mértékegységgel rendelkeznek, szükséges egységes rendszerben történő **újraskálázásuk**, pl. pontérték megadásával (1-5, 1-10, 1-100, stb.). A pontértékek intervallumának megadása már önmagában véve az adatok **manipulációját** jelenti, ráadásul az új skála nominális vagy ordinális lesz.

Az összegzésnél további problémák merülhetnek fel: az adott területegységre jellemző e 3 értékből készített feltérkép végeredménye más lesz, ha összeadjuk és más, ha szorozzuk az egyes változókra jellemző értékeket (ráadásul nominális-ordinális változók esetében nem is végezhetnénk algebrai műveleteket). Mi több, ez az értékek összes statisztikai paraméterét (átlag, módusz stb.) is módosítja. Hasonló módszertani problémákat vet fel a súlyozás kérdése: egyes változókat többszörös súllyal szoktak figyelembe venni bizonyos összevont értékek kiszámításakor, de a súlyfaktor értéke sokszor szubjektív!

Általánosságban azt a módszert érdemes választani, amely segítségével utólag könnyebben besorolhatók az adatok intervallumokba, tehát jobban szét húzza az adatsort, s „hiátusokat” teremt a folytonos eloszlásban.

Általánosságban intervallum-képzésnél azt a módszert érdemes választani, amely segítségével könnyebben besorolhatók az adatok intervallumokba, tehát jobban szét húzza az adatsort, s „hiátusokat” teremt a folytonos eloszlásban.

Ezzel átléptünk az adat-**transzformációk** kérdéskörébe.: Az itt említendő módszereket akkor alkalmazzuk, ha:

**1, eltérő léptékű, nagyságrendű adatok** szerepelnek: ebben az esetben ugyanis bizonyos műveletek (pl. klaszteranalízis) nem végezhetőek el. Ennek megoldásában segít a maximumra vetítés, normalizálás, standardizálás, sorrendi skála alkalmazása

**2, eltérő az adatsorok szórása és terjedelme:** ekkor ugyanis egyes műveletek (pl. 2 mintás t-próba) nem végezhetőek el. Ennek megoldásában segít pl. a normalizálás, de pl. a maximumra vetítés nem (ui. nem húzza szét vagy össze az adatsort).

**a, sorrendiségi skála** alkalmazása (a legnagyobb v. legkisebb érték = 1, a második legnagyobb v. legkisebb = 2), ordinális változókat eredményez, tehát az eredeti metrikus arányskálánk elvész.

**b, a maximumra vetítés** (a legnagyobb érték = 100%, a többi arányosan kisebb), olykor nevezik normálásnak is,

**c, normalizálás és standardizálás:** olyan transzformációk használata, melyek normális eloszlásúvá teszik az adatsort:  $z_i = (x_i - x_{\min}) / (x_{\max} - x_{\min})$ ;  $z_i = \lg(x_i + 1)$ .

Ekkor a statisztikai paraméterek (átlag, módusz, medián, szórás, szélsőérték) értékei megváltoznak, de egymáshoz viszonyított helyzetük (pl. az átlaghoz legközelebb eső adat, a módusz, a medián) megmarad: minden  $x_i, y_i$  adatpárra igaz, hogyha  $x_i < x_j$ , akkor  $z_i < z_j$ .

## 5. 5. Mértékegységek és átszámításuk

A különböző országokban és az eltérő időpontokban készült térképek használatakor mindig figyelniük kell a térkép készítésénél használt mértékegységekre. Ha összehasonlító elemzéseket végzünk, akkor ismernünk kell az átváltás mérőszámát. Legtöbbször a távolság és a terület meghatározása szükséges.

Sok esetben nagy a bizonytalansági tényező, így pl. Fényes Elek egyik munkájában (1836) magyar holdban adja meg a szántóterületek nagyságát, amelyből viszont létezik kis és nagy magyar hold, 1200 ill. 1600 négyszögöl nagyságban. Az 1865-ös adatok kataszteri holdban vannak megadva, manapság viszont hektárt használunk. Így egy térség területhasználat-változásának vizsgálata nem kis kihívást jelent. A legfontosabb táblázatok a mellékletekben találhatóak.

### *Példa:*

Az alábbi paraméterekkel jellemezhető szimmetrikus csatornában víz mozog. Meghatározzuk a víz sebességét a keresztmetszvény jelölt pontjain. Ezek rendre  $v_1=v_6=0,05$  m/s;  $v_2=v_4=0,2$  m/s;  $v_3=v_5=0,1$  m/s. Becsülje meg a csatornában folyó víz átlagsebességét!

Miután a mérési adataink végesek, feltesszük, hogy a mért értékek a víztest egy-egy nagyobb felületének átlagértékei, ezeket az ábrán szaggatott vonallal határoltuk el. Ha ismertek a sebességekhez tartozó keresztmetszetek, akkor a középsebesség egyszerű *súlyozott átlaggal* kiszámolható:

$$\bar{v} = \frac{\sum v_i A_i}{\sum A_i}$$

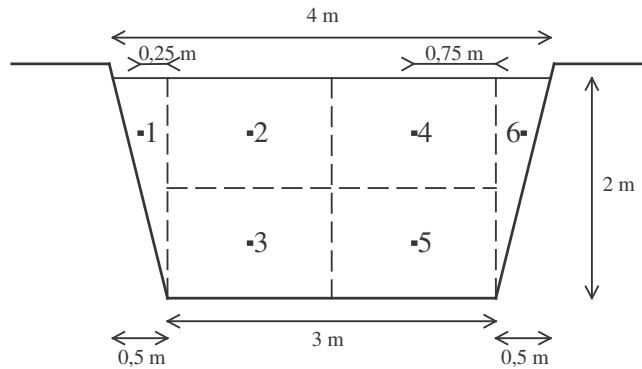
Az 1. és 6. háromszög egybevágó, derékszögű, területük:

$$A_1 = A_6 = \frac{0,5 \cdot 2}{2} \text{ m}^2 = 0,5 \text{ m}^2$$

A 2.–5. téglalap területe:

$$A_2 = A_3 = A_4 = A_5 = 1 \cdot 1,5 \text{ m}^2 = 1,5 \text{ m}^2$$

$$\begin{aligned} \bar{v} &= \frac{\sum v_i A_i}{\sum A_i} = \frac{v_1 A_1 + v_2 A_2 + v_3 A_3 + v_4 A_4 + v_5 A_5 + v_6 A_6}{A_1 + A_2 + A_3 + A_4 + A_5 + A_6} = \\ &= \frac{0,05 \cdot 0,5 + 0,2 \cdot 1,5 + 0,1 \cdot 1,5 + 0,2 \cdot 1,5 + 0,1 \cdot 1,5 + 0,05 \cdot 0,5}{0,5 + 1,5 + 1,5 + 1,5 + 1,5 + 0,5} \frac{\text{m}}{\text{s}} = 0,136 \frac{\text{m}}{\text{s}} \end{aligned}$$



*Megjegyzés:* A  $Q = \sum v_i A_i$  képlet a vízhozamot adja meg. A számolással becslés jellegű vízhozamot adunk meg, mert egyrészt nem teljesül feltétlenül a sebességvektor és a felület merőlegessége, másrészt a felületekhez rendelt, mért sebességértékek a középsebességtől eltérnek.

### Feladatok

1, Egy 5,4 hektáros parcella terméseredményei (t) láthatók lent, a bal oldali táblázatban. Számolja ki az átlagos hozamot, a szórást, móduszt, mediánt. A többi ábrán mintavételezési típusokat látunk (x a minta helye). Nevezze meg a mintavételezés típusát, majd számolja ki a mintavételezéssel kapott termésátlagot (szórást, stb.) és vesse össze a teljes terület paramétereivel.

1	2	2	2
1	1	2	2
2	2	2	2
2	2	3	0
2	1	1	4

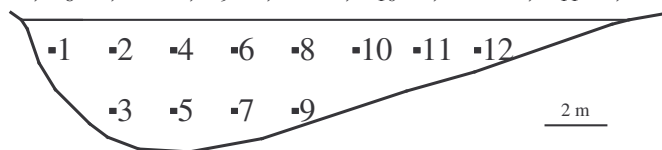
x	x	x	
	x		
		x	x
		x	x

x		x	
	x		x
x		x	
	x		x
x		x	

x		x	
	x	x	
x			x
		x	
x	x		

- Hány km<sup>2</sup>, ár, m<sup>2</sup>, kataszteri hold, kis magyar hold a fenti 20 hektáros parcella?
- Végezzük el az alábbi folyószelvény középsebességének becslés jellegű meghatározását, ha rendelkezésünkre áll az alábbi mederkeresztmetszet-rajz és a jelölt pontokon mért sebességadatok!

$v_1=0,1$  m/s;  $v_2=0,15$  m/s;  $v_3=0,1$  m/s;  $v_4=0,2$  m/s;  $v_5=0,15$  m/s;  $v_6=0,2$  m/s;  $v_7=0,15$  m/s;  $v_8=0,2$  m/s;  $v_9=0,1$  m/s;  $v_{10}=0,15$  m/s;  $v_{11}=0,12$  m/s;  $v_{12}=0,1$





## 6. Mátrixok. Mátrixműveletek és tulajdonságaik

A gyakorlati életben adatok gyűjtésénél, műszeres mérések során a számokat sokszor táblázatszerű formába rendezzük, mert így a sokoldalú összefüggések áttekinthetőbbek. A számok ilyen táblázatszerű elrendezését **mátrixnak** nevezzük.

A mátrix vízszintes vonalban elhelyezkedő elemeit **soroknak**, függőleges vonalban elhelyezkedő elemeit pedig **oszlopoknak** nevezzük. Egy  $n$  sorból és  $m$  oszlopból álló mátrixot  $n$ -szer  $m$  ( $n \times m$ ) mátrixnak nevezzük. A mátrixokat nagybetűvel (pl.  $A$ ), a mátrix elemeit pedig kisbetűvel (pl.  $a$ ) jelöljük.

Az  $A$  mátrix jelölése:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1m} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ a_{i1} & \dots & a_{ik} & \dots & a_{im} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nm} \end{pmatrix} \text{ vagy } \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1m} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ a_{i1} & \dots & a_{ik} & \dots & a_{im} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nm} \end{bmatrix}$$

Beszélhetünk a mátrix  $i$ -edik soráról és  $k$ -edik oszlopáról. A mátrixnak az  $i$ -ik sorban és  $k$ -ik oszlopban lévő elemét a mátrix  $ik$ -edik elemének nevezik, jelölése  $a_{ik}$ . Mindig először a sorszám, majd az oszlopszám szerepel. Általában a mátrix sorainak és oszlopainak számozása 1-gyel kezdődik, de vannak olyan számítógépes programok, melyek 0-val kezdenek. A fenti  $A$  mátrix téglalap alakú, de nemcsak téglalap alakú mátrixok vannak. Az olyan mátrixot, amelynél a sorok és az oszlopok száma megegyezik ( $n=m$ ), négyzetes, vagy kvadratikus mátrixnak nevezzük. Az olyan négyzetes mátrixot, melynek csak főátlójában vannak 0-tól eltérő elemek diagonálmátrixnak nevezzük. Például egy harmadrangú ( $n=3$ ) diagonál mátrix:

$$\begin{pmatrix} 3 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Egy mátrixot akkor tekintünk ismertnek, ha tudjuk, hogy milyen alakú és ismerjük a mátrix elemeit alkotó számokat.

Az olyan mátrixot, amelynek csak egy sora, vagy egy oszlopa van **vektornak** (többdimenziós vektornak) hívjuk. Így a mátrixok felfoghatók a vektorok általánosításának is. Az egy sorú és egy oszlopú mátrix pedig egy skaláris mennyiség.

A **sorvektornak** csak egy sora van:

$$(a_1 \quad a_2 \quad a_3 \quad \dots \quad a_m)$$

Az **oszlopvektornak** pedig egyetlen oszlopa van:

$$\begin{pmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{pmatrix}$$

Ha egy mátrix sorait az oszlopaival felcseréljük, akkor általában más mátrixot kapunk. Ezt az új mátrixot a régi mátrix transzponáltjának nevezzük. Az adatbázisok felfoghatók mátrixként, előfordulhat, hogy az oszlopokból (változók), sorokat (esetek) kell csinálni (Excel, SPSS), ezért a transzponálás fontos művelet.

A fenti  $A$  mátrix transzponáltja:

$$A^T = \begin{pmatrix} a_{11} & a_{21} & a_{31} & \dots & a_{n1} \\ a_{12} & a_{22} & a_{32} & \dots & a_{n2} \\ \dots & \dots & \dots & \dots & \dots \\ a_{1i} & \dots & a_{ki} & \dots & a_{ni} \\ \dots & \dots & \dots & \dots & \dots \\ a_{1m} & a_{2m} & a_{3m} & \dots & a_{nm} \end{pmatrix}$$

A nullmátrixban minden  $a_{ik}$  elem értéke 0. A négyzetes mátrixot akkor nevezzük  $E$  egységmátrixnak, ha diagonálisan, vagyis azokon a helyeken, ahol  $i=k$ , mindig az 1 szám áll, és minden más  $i \neq k$  helyen 0 van. A (3, 3) egységmátrix pl.:

$$E = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Két mátrix akkor egyenlő, ha mindegyiknek ugyanannyi sora és ugyanannyi oszlopa van és a megfelelő helyen álló elemek rendre megegyeznek, azaz  $A = B$ , ha minden  $i$ -re és  $k$ -ra  $a_{ik} = b_{ik}$ .

A négyzetes  $n \times n$ -es  $A$  mátrix inverzén értjük azt a négyzetes  $A^{-1}$  mátrixot, amelyre teljesül, hogy  $A \cdot A^{-1} = E$

Az  $n \times n$ -es mátrix lehet **invertálható**, **reguláris** vagy **szinguláris**. Egy  $n$ -szer  $n$ -es  $A$  mátrix akkor és csakis akkor invertálható, ha létezik egy olyan  $B$  mät-

rix, melyre igaz:  $\mathbf{AB} = \mathbf{I}_n (= \mathbf{BA})$ . Ebben az esetben a  $\mathbf{B}$  mátrix az  $\mathbf{A}$  mátrix **inverz mátrixa** és  $\mathbf{A}^{-1}$ -al jelölik. Az  $\mathbf{I}_n$  az  $n$ -szer  $n$ -es egységmátrixot jelöli és a szorzás a szokásos mátrixszorzás. Ha a mátrix nem invertálható, akkor **szingulárisnak** nevezzük.

## 6. 1. Műveletek mátrixokkal

### Összeadás - kivonás

Összeadni és kivonni csak azonos típusú mátrixokat lehet. Az összegmátrix elemei ekkor az összeadandó mátrixok megfelelő elemeinek összegei. Legyen  $\mathbf{A}$  és  $\mathbf{B}$  két  $n$ -szer  $m$  méretű mátrix, akkor  $\mathbf{C} = \mathbf{A} + \mathbf{B}$ , illetve  $\mathbf{D} = \mathbf{A} - \mathbf{B}$ , ha

$$c_{ik} = a_{ik} + b_{ik}, \text{ illetve } d_{ik} = a_{ik} - b_{ik} \text{ minden } i\text{-re és } k\text{-ra.}$$

Ha pl:

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 0 \\ -1 & 1 & 3 \end{pmatrix} \text{ és } \mathbf{B} = \begin{pmatrix} -1 & 0 & 2 \\ 3 & 1 & 2 \end{pmatrix},$$

$$\text{akkor } \mathbf{C} = \mathbf{A} + \mathbf{B} = \begin{pmatrix} 0 & 2 & 2 \\ 2 & 2 & 5 \end{pmatrix} \text{ és } \mathbf{D} = \mathbf{A} - \mathbf{B} = \begin{pmatrix} 2 & 2 & -2 \\ -4 & 0 & 1 \end{pmatrix},$$

$$\text{mert pl. } c_{23} = a_{23} + b_{23} = 3 + 2 = 5, \text{ és } d_{23} = a_{23} - b_{23} = 3 - 2 = 1.$$

### Skalárral való szorzás

Egy  $\mathbf{A}$  mátrixot valamely  $k$  számmal (skalárral) úgy szorzunk, hogy a mátrix minden elemét megszorozzuk  $k$ -val.  $\mathbf{C} = k \cdot \mathbf{A}$  azt jelenti, hogy  $c_{ik} = k \cdot a_{ik}$ .

$$\text{Ha } k = 3 \text{ és } \mathbf{A} = \begin{pmatrix} 1 & 2 & 0 \\ -1 & 1 & 3 \end{pmatrix} \text{ } 3\mathbf{A} = \begin{pmatrix} 3 & 6 & 0 \\ -3 & 3 & 9 \end{pmatrix}$$

### Mátrixok szorzása

A szorzatmátrix elemeit úgy definiáljuk, mint az első mátrix egy sorvektorának a második mátrix egy oszlopvektorával való skaláris szorzatát.

Ha egy  $\mathbf{a} = (a_1, \dots, a_n)$  sorvektort megszorozunk egy  $\mathbf{b} = \begin{pmatrix} b_1 \\ \cdot \\ b_n \end{pmatrix}$  oszlopvektorral.

Ezek skaláris szorzatát a következőképpen határozzuk meg:

$$\mathbf{a} \cdot \mathbf{b} = a_1b_1 + a_2b_2 + \dots + a_nb_n. \text{ Ha pl. } \mathbf{a} = (3 \ 1 \ -2) \text{ és } \mathbf{b} = \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix},$$

$$\text{akkor } \mathbf{a} \cdot \mathbf{b} = 3 \cdot 3 + 1 \cdot (-1) + (-2) \cdot 2 = 4$$

$$\text{Legyen } \mathbf{A} = \begin{pmatrix} 1 & 0 & 3 & 1 \\ 2 & 1 & 0 & 1 \\ 1 & 2 & 3 & 2 \end{pmatrix} \text{ és } \mathbf{B} = \begin{pmatrix} 2 & 0 & 1 \\ 3 & 1 & 3 \\ 0 & 2 & 2 \\ 1 & 0 & 2 \end{pmatrix} \text{ és határozzuk meg a } \mathbf{A} \cdot \mathbf{B} = \mathbf{C} \text{ -t}$$

$$\mathbf{C} = \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix}$$

A szorzás definiálása alapján:

$$c_{11}=1 \cdot 2 + 0 \cdot 3 + 3 \cdot 0 + 1 \cdot 1 = 3 \quad c_{12}=1 \cdot 0 + 0 \cdot 1 + 3 \cdot 2 + 1 \cdot 0 = 6 \quad c_{13}=1 \cdot 1 + 0 \cdot 3 + 3 \cdot 2 + 1 \cdot 2 = 9$$

$$c_{21}=2 \cdot 2 + 1 \cdot 3 + 0 \cdot 0 + 1 \cdot 1 = 8 \quad c_{22}=2 \cdot 0 + 1 \cdot 1 + 0 \cdot 2 + 1 \cdot 0 = 1 \quad c_{23}=2 \cdot 1 + 1 \cdot 3 + 0 \cdot 2 + 1 \cdot 2 = 7$$

$$c_{31}=1 \cdot 2 + 2 \cdot 3 + 3 \cdot 0 + 2 \cdot 1 = 10 \quad c_{32}=1 \cdot 0 + 2 \cdot 1 + 3 \cdot 2 + 2 \cdot 0 = 8 \quad c_{33}=1 \cdot 1 + 2 \cdot 3 + 3 \cdot 2 + 2 \cdot 2 = 17$$

$$\text{Ennek megfelelően a } \mathbf{C} = \begin{pmatrix} 3 & 6 & 9 \\ 8 & 1 & 7 \\ 10 & 8 & 17 \end{pmatrix}$$

A fenti példa alapján az  $m \times n$  típusú  $\mathbf{A}$  és  $n \times k$  típusú  $\mathbf{B}$  mátrixok szorzatán azt az  $m \times k$  típusú  $\mathbf{C}$  mátrixot értjük, amelynek elemeit a következő képlettel határozhatjuk meg:

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}, \text{ ahol } i=1, \dots, m \text{ és } j=1, \dots, k.$$

**A szorzás definíciója csak olyan mátrixokra vonatkozik, ahol az első mátrixnak annyi sora van, ahány oszlopa a másíknak.**

## 6. 2. Mátrixműveletek tulajdonságai

Az összeadás tulajdonságai:

kommutativitás

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$$

asszociativitás

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$$

disztributivitás

$$k(\mathbf{A} + \mathbf{B}) = k\mathbf{A} + k\mathbf{B} \text{ és } (k+l)\mathbf{A} = k\mathbf{A} + l\mathbf{A}$$

A mátrixok szorzásának a tulajdonságai:

asszociativitás

$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$  minden  $k$ -szor  $m$  méretű  $\mathbf{A}$  mátrixra,  $m$ -szernméretű  $\mathbf{B}$  mátrixra és  $n$ -szer  $p$  méretű  $\mathbf{C}$  mátrixra.

disztributivitás

$(\mathbf{B} + \mathbf{C})\mathbf{A} = \mathbf{BA} + \mathbf{CA}$  minden  $m$ -szer  $n$  méretű  $\mathbf{A}$  és  $\mathbf{B}$  mátrixra valamint  $n$ -szer  $k$  méretű  $\mathbf{C}$  mátrixra, valamint

$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$  minden  $m$ -szer  $n$  méretű  $\mathbf{A}$  és  $\mathbf{B}$  valamint  $k$ -szor  $m$  méretű  $\mathbf{C}$  mátrixra.

kommutativitás

nem teljesül; vagyis adott  $\mathbf{A}$  és  $\mathbf{B}$  összeszorozható mátrixra általában igaz, hogy  $\mathbf{AB} \neq \mathbf{BA}$ .

$$(\lambda \cdot \mathbf{A}) \cdot \mathbf{B} = \mathbf{A} \cdot (\lambda \cdot \mathbf{B}) = \lambda \cdot (\mathbf{A} \cdot \mathbf{B})$$

### ***Feladatok***

Mikor egyenlő a következő két mátrix?

$$\mathbf{A} = \begin{pmatrix} 5 & 7 & 9 \\ 2 & 8 & 6 \\ 4 & 3 & 7 \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} a & 7 & 9 \\ 2 & b+3 & 6 \\ c-2 & 3 & 7 \end{pmatrix}$$

Mátrixok összeadása, kivonása

a,  $\begin{pmatrix} -2 & 6 \\ 3 & 5 \end{pmatrix} + \begin{pmatrix} 4 & 3 \\ 7 & -1 \end{pmatrix} =$

b,  $\begin{pmatrix} 3 & 6 \\ 4 & -2 \end{pmatrix} - \begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix} =$

c,  $\begin{pmatrix} 0 & 1 & 3 \\ 2 & 4 & 6 \end{pmatrix} + \begin{pmatrix} 2 & -3 \\ 5 & 1 \\ 0 & -2 \end{pmatrix} =$

d,  $\begin{pmatrix} 3 & 6 & -7 \\ 2 & 1 & 4 \end{pmatrix} - \begin{pmatrix} 2 & 2 & -2 \\ -1 & 0 & 3 \end{pmatrix} =$

Mátrixok szorzása

$$\text{a, } \begin{pmatrix} -2 & 6 \\ 3 & 5 \end{pmatrix} \cdot \begin{pmatrix} 4 & 3 \\ 7 & -1 \end{pmatrix} =$$

$$\text{b, } \begin{pmatrix} 3 & 6 \\ 4 & -2 \end{pmatrix} \cdot \begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix} =$$

$$\text{c, } \begin{pmatrix} 0 & 1 & 3 \\ 2 & 4 & 6 \end{pmatrix} \cdot \begin{pmatrix} 2 & -3 \\ 5 & 1 \\ 0 & -2 \end{pmatrix} =$$

$$\text{d, } \begin{pmatrix} 3 & 6 & -7 \\ 2 & 1 & 4 \end{pmatrix} \cdot \begin{pmatrix} 2 & 2 & -2 \\ -1 & 0 & 3 \end{pmatrix} =$$

$$\text{e, } (4 \ 2 \ 3) \cdot \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix} =$$

$$\text{f, } \begin{pmatrix} 1 \\ -3 \\ 2 \end{pmatrix} \cdot (4 \ 3 \ 2) =$$

$$\text{g, } \begin{pmatrix} 6 & 7 & 3 \\ 3 & 1 & 0 \end{pmatrix} \cdot (4 \ 3 \ 2) =$$

$$\text{h, } \left( \begin{bmatrix} 3 & 4 \\ 1 & -2 \end{bmatrix} \cdot \begin{bmatrix} 2 & -3 \\ 1 & 5 \end{bmatrix} \right) \cdot \begin{pmatrix} 1 & 3 \\ 3 & 8 \end{pmatrix} =$$

$$\text{i, } \left( \begin{bmatrix} 3 & 4 \\ 1 & -2 \end{bmatrix} + \begin{bmatrix} 2 & -3 \\ 1 & 5 \end{bmatrix} \right) \cdot \begin{pmatrix} 1 & 3 \\ 3 & 8 \end{pmatrix} = \text{ j, } \begin{pmatrix} 1 & 4 \\ 3 & 3 \end{pmatrix} \cdot \begin{pmatrix} -1 & 2 \\ 2 & 3 \end{pmatrix} \cdot \begin{pmatrix} 1 & 3 \\ 4 & 8 \end{pmatrix} \cdot \begin{pmatrix} -2 & 2 \\ 3 & 4 \end{pmatrix} =$$

$$\text{k, } 4 \cdot (3 \ 2 \ 1) =$$

$$\text{l, } 3 \cdot \begin{pmatrix} 4 & 3 & 2 \\ 2 & -1 & 4 \end{pmatrix}$$

## 7. Halmazok, halmazműveletek és tulajdonságaik

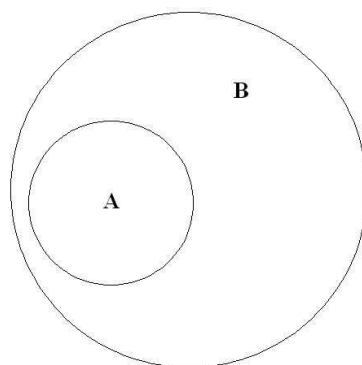
### 7.1. A halmaz fogalma

"A halmaz érzékelésünk és gondolkodásunk jól meghatározott és egymástól megkülönböztethető tárgyainak egységbe foglalása." (Georg Cantor 1845-1918). A matematikában nem meghatározandó alapfogalomnak tekintik. A halmazba foglalt dolgokat a halmaz elemeinek nevezzük, tehát úgy is mondhatjuk, hogy a halmazokat az elemeik határozzák meg. (Nincs korlátozva a halmazok elemeinek a száma.) Egy halmazt végesnek mondunk, ha véges sok eleme van, tehát az elemeinek a számát egy természetes számmal meg lehet adni (pl. az év hónapjai). A halmazt végtelennek nevezzük, ha végtelen sok eleme van (pl. természetes számok halmaza). Azt a halmazt, amelynek nincs egyetlen eleme sem, üres halmaznak nevezzük.

A halmazok jelölésére a nagybetűket használjuk. Ha „a” egy H halmaz eleme, akkor ezt röviden a következő módon jelöljük:  $a \in H$ , abban az esetben, ha „a” nem eleme H-nak, akkor  $a \notin H$  meg lehet adni.

Ha két halmaz elemei megegyeznek, akkor a két halmazt azonosnak nevezzük. Egy halmazban egy dolog csak egyszer szerepelhet elemként (pl. az év hónapjai között nem szerepelhet kétszer a január). A halmaz elemeit kapcsos zárójelek  $\{ \{1, 2, 5\} \}$  közé írjuk. Az üres halmaz jele:  $\emptyset$

Egy A halmaz részhalmaza (része) B halmaznak, ha A-nak minden eleme része a B-nek. Jelölése:  $A \subseteq B$



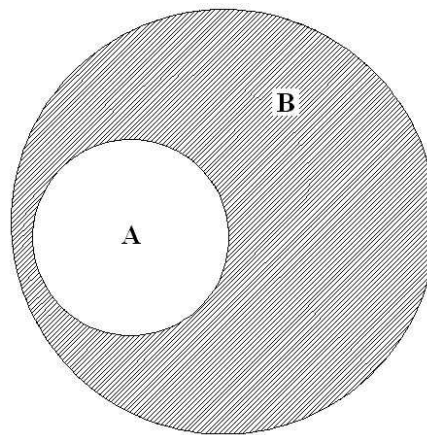
A hétköznapi életben a „rész” fogalma kevesebbet jelent, mint az „egész”, viszont a halmazelméletben az „egész” is beletartozik a „rész” fogalmába.) Ennek megfelelően, minden halmaz része önmagának ( $A \subseteq A$ ). Ezt reflexivitás tulajdonságnak nevezzük. Az üres halmaz pedig minden halmaznak a része.

Egy  $A$  halmaz valódi részhalmaza  $B$  halmaznak, ha  $A$  része  $B$ -nek, de nem azonos vele. Jelölése:  $A \subset B$

A „ $\subset$ ” reláció tulajdonságai:

- ha  $A \subset B$ , akkor  $A \subseteq B$  (a valódi rész mindig rész is)
- a  $A \subset A$  sohasem igaz (a „ $\subset$ ” irreflexivitása)
- ha  $A \subset B$ , akkor  $B \not\subset A$  (a „ $\subset$ ” aszimmetriája)
- ha  $A \subset B$ , és  $B \subset C$  akkor  $A \subset C$  (a „ $\subset$ ” tranzitivitása)
- ha  $A \subseteq B$ , és  $B \subseteq C$  akkor  $A \subseteq C$  (a „ $\subseteq$ ” tranzitivitása)

Ha egy  $A$  halmaz részhalmaza egy  $B$  halmaznak, akkor azt a halmazt, amely a  $B$  minden olyan elemét tartalmazza, amely nem eleme  $A$ -nak, az  $A$  halmaz  $B$  halmazra vonatkozó komplementerének nevezzük.



- egy halmaznak önmagára vonatkoztatott komplementuma az üres halmaz
- az üres halmaz komplementuma maga az alaphalmaz

## 7. 2. Halmazműveletek és tulajdonságaik

- **Halmazok metszete**

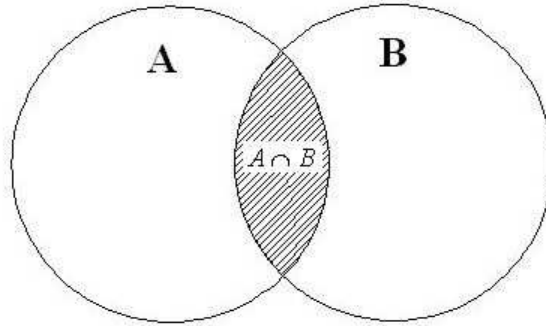
Két halmaz,  $A$  és  $B$  metszetének nevezzük azoknak az elemeknek  $C$  halmazát, amelyek  $A$ -nak is és  $B$ -nek is elemei.

A metszet tehát a közös elemek halmaza.

$$\text{Jelölése: } A \cap B = C = \{x \mid x \in A \text{ és } x \in B\}$$

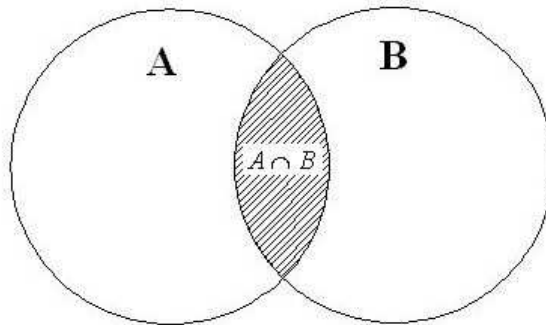
$A$ -t és  $B$ -t **diszjunkt** (idegen) halmazoknak nevezzük, ha nincs közös elemük, azaz metszetük üres halmaz.



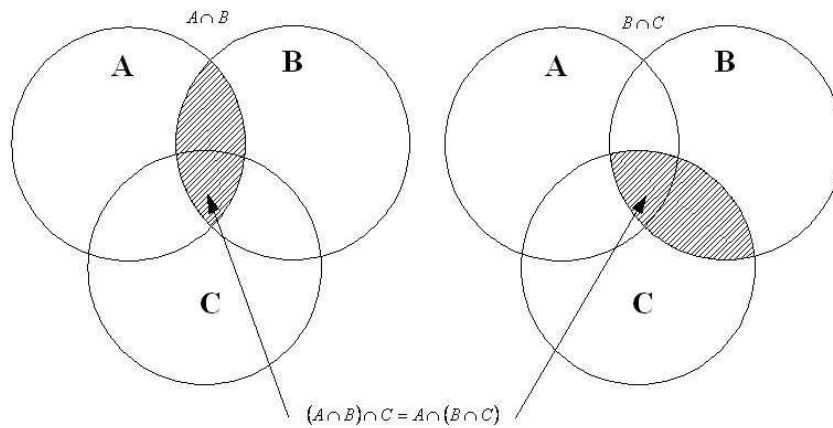


**Tulajdonságok:**

- $A \cap B = B \cap A$ , azaz **kommutatív**
- Három halmaz metszete **asszociatív**:  $(A \cap B) \cap C = A \cap (B \cap C)$



- A metszet az unióra nézve **disztributív**:  
 $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ .



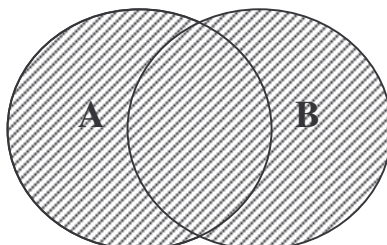
- Egy halmaz üres halmazzal képzett metszete az üres halmaz.  
 $H \cap \emptyset = \emptyset$ .

- Egy halmaznak a komplementumával képzett metszete ugyancsak üres halmaz.  $H \cap H' = \emptyset$ .

- **Halmazok uniója**

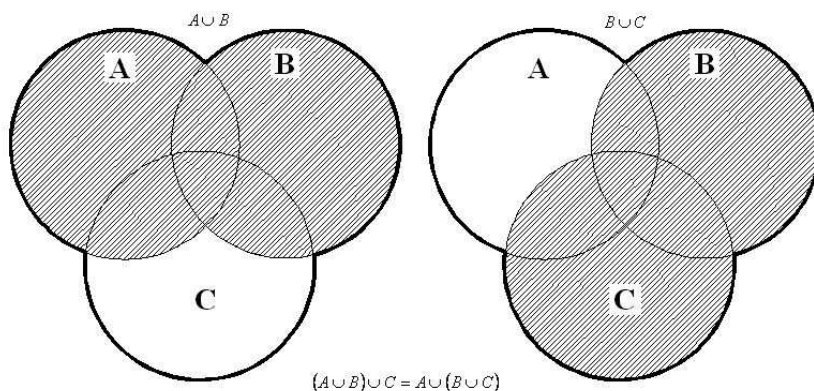
Az  $A$  és  $B$  halmazok uniójának (egyesített halmazának) nevezzük azt a halmazt, amelynek elemei az  $A$  és  $B$  halmazok közül legalább az egyiknek elemei. Jele:  $A \cup B = \{x \in A \text{ vagy } x \in B\}$

(Az  $x$  lehet  $A$ -nak is és  $B$ -nek is eleme)



**Tulajdonságok:**

- $A \cup B = B \cup A$ , azaz **kommutatív**
- Három halmaz uniója **asszociatív**:  $(A \cup B) \cup C = A \cup (B \cup C)$



- $A \cup A = A$
- Az üres halmazzal képzett unió maga a halmaz:  $A \cup \emptyset = A$
- Az unió a metszetre nézve disztributív :  

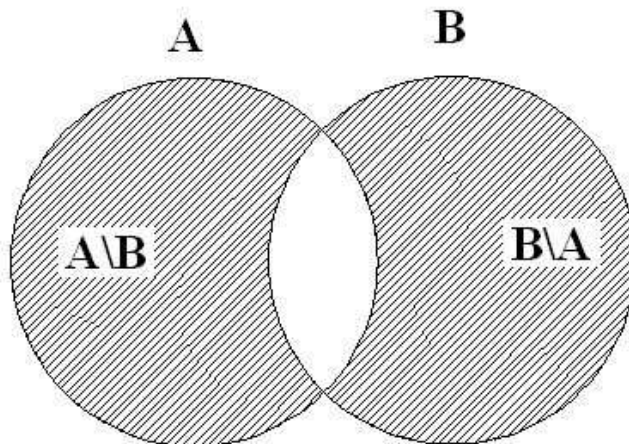
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

- **Halmazok különbsége**

$A$  és  $B$  halmazok különbsége  $A$  halmaz azon elemeinek halmaza, amelyek nem elemei  $B$ -nek. Jelölése:  $A \setminus B = \{x \mid x \in A \text{ és } x \notin B\}$

**Tulajdonságok:**

- Nem kommutatív:  $A \setminus B \neq B \setminus A$
- Nem asszociatív:  $(A \setminus B) \setminus C \neq A \setminus (B \setminus C)$

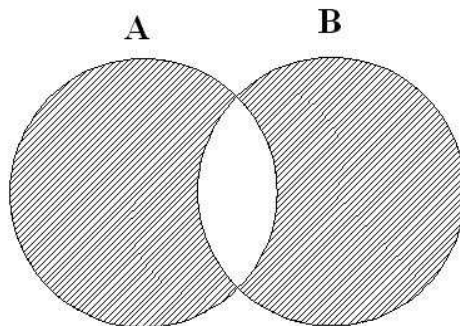


- Az üres halmaznak nincs eleme, ezért nyilvánvaló, hogy  
 $A \setminus \emptyset = A$ ,  $\emptyset \setminus A = \emptyset$  és  $A \setminus A = \emptyset$ .
  - $(A \setminus B) \cap (B \setminus A) = \emptyset$
  - $(A \setminus B) \cup B = A \cup B$
  - $(A \setminus B) \cap A = A \setminus B$
  - $(A \setminus B) \cap B = \emptyset$
- **Szimmetrikus különbség**

Két halmaz szimmetrikus különbsége:  $(A \setminus B) \cup (B \setminus A)$

Azaz  $A$  és  $B$  halmazok szimmetrikus különbsége azoknak az elemeknek a halmaza, amelyek elemei vagy  $A$ -nak, vagy  $B$ -nek, de nem mindkettőnek.

Jelölése:  $A \Delta B$ , ahol  $A$  és  $B$  a szimmetrikus különbség tagjai.



### Feladatok

1. Adott három halmaz:  $A = \{a, b, c, d, e, f\}$ ,  $B = \{b, d, e\}$  és  $C = \{c, f\}$ .  
Határozza meg a következő halmazok elemeit!
  - a,  $A \cap B$  megoldás:  $\{b, e\}$
  - b,  $A \cup B$  megoldás:  $\{a, b, c, d, e, f\}$
  - c,  $A \cap C$  megoldás:  $\{c, f\}$
  - d,  $B \cup C$  megoldás:  $\{b, c, d, e, f\}$
  - e,  $B \cap C$  megoldás:  $\emptyset$
  
2. Legyen  $A = \{\text{Pécs, Debrecen, Békés, Eger, Győr, Szeged, Pest}\}$  és  $B = \{\text{Vas, Zala, Baranya, Pécs, Pest, Békés, Győr}\}$  két halmaz. Határozza meg uniójukat, metszetüket, különbségüket, szimmetrikus differenciájukat. Képezzen belőlük legalább három-három részhalmazt!
  
3. Legyen a vizsgálatunk alaphalmaza (H) a Föld országai. Értelmezzük a következő halmazokat az alaphalmazon belül:  $A = \{\text{EU tagországok}\}$ ,  $B = \{\text{Schengeni Egyezményt aláíró országok}\}$ ,  $C = \{\text{euró-zóna tagországai}\}$ ,  $D = \{\text{az ENSZ Biztonsági Tanács tagországai}\}$ . Végezze el az alábbi halmazműveleteket, és diszkutálja az eredményt!
  - a,  $A \cap (H \setminus B)$  (azok az országok, melyek az EU tagjai és nem írták alá a Schengeni Egyezményt)
  - b,  $C \setminus A$  (azon országok, melyek nem az EU tagjai, de a hivatalos fizetőeszközük az euró)Fejtse ki, hogy mit jelentenek a következő műveletek a fent definiált halmazokon!
  - c,  $(A \setminus B) \cup (B \setminus A)$
  - d,  $A \cup B \cup C$
  - e,  $A \cap D$

## 8. Valószínűségszámítás

A következő két fejezetben a valószínűségszámítás témakörével foglalkozunk. Természetesen itt sem tűzhetjük ki célul a valószínűségszámítás elméleti és gyakorlati vonatkozású témaköreinek teljes áttekintését, csak a geográfiai alkalmazásoknál gyakran előforduló tételeket emeljük ki. Ehhez azonban nélkülözhetetlen az alapfogalmak és néhány kapcsolódó témakör megismerése.

A valószínűségszámítás feladata olyan mérték bevezetése, amely a bizonytalanságot numerikusan méri és erre alapozva olyan matematikai módszereket dolgoz ki, amelyekkel bizonyos események (véletlen tömegjelenségek) modellezhetők, illetve a valószínűségek kiszámíthatók.

### 8.1. A valószínűségszámítás alapjai

A vizsgálataink során alapvetően kétféle jelenséggel találkozhatunk. Az egyik az, amikor előre meg tudjuk mondani, hogy mi fog történni. Például, ha egy vízfolyás áradásakor a vízszint megközelíti a gát tetejét és újabb jelentős vízutánpótlás várható, akkor biztos átszakítja a gátat és elönti a nem mentett árteret. Ezt a bekövetkezett árvizet egy „eseménynek” nevezzük. Másik esemény például, hogy nincs vízutánpótlás, nem emelkedik tovább a vízszint. (Ez ebben az esetben nem következik be az árvíz.) Az ilyen jelenségeket nevezzük *determinisztikus jelenségnek*.

A másik fajta jelenség például az, ha nem tudjuk előre megmondani, hogy a sok esemény közül melyik fog bekövetkezni (Pl. vihar várható, de nem tudjuk, hogy milyen károkat idéz elő). Tehát, ha nem ismerjük eléggé a körülményeket, a feltételeket, akkor nem tudjuk előre megmondani, mi fog történni. Az ilyen jelenségeket hívjuk *véletlen jelenségeknek*.

Vannak olyan jelenségek, amiket sokszor meg tudunk figyelni (pl. a szélsebesség és talajszállítás közötti összefüggés). Egy ilyen megfigyelést szokták *„kísérletnek”* is nevezni.

Egy kísérlet kimenetelére különböző állításokat fogalmazhatunk meg. Ha az állítás igaz vagy hamis volta csak a kísérlet kimenetelétől függ, akkor az állítást *eseménynek* nevezzük. Jelölése általában nagybetűvel történik.

#### *Alapfogalmak*

- elemi esemény  $(k_i)$  az  $i$ -edik kísérlet eredménye
- eseménytér  $\mathbf{H}=\{k_1, k_2, \dots, k_n\}$  az elemi események halmaza
- összetett esemény  $A \subseteq \mathbf{H}$  az eseménytér részhalmaza
- lehetetlen esemény – amely sohasem következhet be:  $\emptyset$

- biztos esemény – amely a kísérlet során biztosan bekövetkezik, maga az eseménytér: **H**

## 8. 2. Műveletek eseményekkel

- ***Komplementer esemény***

Az  $A$  esemény komplementere vagy ellentett eseménye a „nem  $A$ ”. Jele:  $\bar{A}$

- ***Események összege***

Események összegén azt az eseményt értjük, ami pontosan akkor következik be, ha az események közül legalább az egyik bekövetkezik.

$A+B$  vagy  $A \cup B$   $k \in A$  **vagy**  $k \in B$  lesz.

- ***Események szorzata***

Két esemény szorzata az az esemény, mely pontosan akkor következik be, ha mindkét esemény bekövetkezik.  $A \cdot B$  vagy  $A \cap B$  eredménye  $k \in A$  **és**  $k \in B$  lesz.

- ***Események különbsége***

$A-B$  vagy  $A \setminus B$  az az esemény, mely akkor következik be, ha a kísérlet eredménye  $k \in A$ , **de**  $k \notin B$  lesz.

$$\text{Fennáll, hogy: } A - B = A \cdot \bar{B}$$

- ***Események azonossága***

$A=B$  ha az  $A \subseteq B$  és  $B \subseteq A$  egyidejűleg teljesül

### Műveleti tulajdonságok

$A+A=A$	$A \cdot A=A$
$A+B=B+A$	$A \cdot B=B \cdot A$
$(A+B)+C=A+(B+C)=A+B+C$	$(A \cdot B) \cdot C=A(B \cdot C)$
$(A+B) \cdot C=A \cdot C+B \cdot C$	$(A \cdot B)+C=A \cdot C+B \cdot C$
$A + \bar{A} = H$	$A \cdot \bar{A} = \emptyset$
$A + \emptyset = A$	$A \cdot \emptyset = \emptyset$
$A+H=H$	$A \cdot H=A$
$\overline{A \cdot B} = \bar{A} + \bar{B}$	$\overline{A + B} = \bar{A} \cdot \bar{B}$
$\bar{\bar{A}} = A$	

Az az esemény, ami soha nem következik be, a **lehetetlen esemény**. Jele:  $\emptyset$

- **Egymást kizáró események**

Legyenek A és B események ugyanannak az eseménytérnek részei. Ha  $A \cdot B = \emptyset$ , akkor A és B események egymást kizáró események.

Legyenek  $A_1, A_2, \dots, A_n$  események részei ugyanannak az eseménytérnek. Ha fennáll, hogy  $A_i \cdot A_j = \emptyset$  minden  $i \neq j$ -re, akkor az  $A_1, A_2, \dots, A_n$  eseményeket egymást páronként kizáró eseményeknek nevezzük.

Az az esemény, ami biztosan bekövetkezik, a biztos esemény. pl.  $x + \bar{x}$  biztos esemény.

Ha egy esemény előáll legalább két másik esemény összegeként, akkor összetett, ha nem, akkor elemi eseménynek nevezzük.

### 8. 3. A valószínűség fogalma

Sokszor találkozunk olyan kísérletekkel, amelyekben az elemi események száma véges és esélyük (valószínűségük) megegyezik – ekkor klasszikus valószínűségi problémáról beszélünk.

Egy kísérletet n-szer elvégezve az A esemény k-szor ( $k \leq n$ ) következik be. Az A esemény relatív gyakorisága:

$$P(A) = \frac{\text{kedvező elemi események száma}}{\text{összes elemi események száma}} = \frac{k}{n}$$

A kísérletek számának növelésével  $P(A)$  értéke stabilizálódik (egyre jobban közelít egy adott értékhez), akkor ezt az értéket nevezzük az  $A$  esemény valószínűségének, amit  $P(A)$  -val jelölünk.

Egy konkrét kísérlet összes lehetséges eseményeihez tartozik egy-egy számérték, amit az illető esemény valószínűségének nevezünk, és amelyre a következő axiómák teljesülnek:

- I. A biztos esemény valószínűsége:  $P(A)=1$
- II. Minden "A" eseményre  $0 \leq P(A) \leq 1$
- III. Ha  $A_1, A_2, \dots, A_n$  események egymást páronként kizáró események, akkor  $P(A_1+A_2+\dots+A_n) = P(A_1)+P(A_2)+\dots+P(A_n)$

Ha az eseményteret alkotó  $n$  számú elemi események ( $E_1, E_2, \dots, E_n$ ) egyformán valószínűek, akkor a hozzájuk rendelt valószínűség:

$$P(E_k) = \frac{1}{n}$$

Az  $A = E_1 + E_2 + \dots + E_k$  összetett esemény valószínűsége:

ha az  $E_i \cap E_j = \emptyset \quad \forall i \neq j$

$$P(A) = P(E_1 + E_2 + \dots + E_k) = P(E_1) + P(E_2) + \dots + P(E_k) = \frac{k}{n}$$

#### 8. 4. Geometriai valószínűség

Ha az eseménytér nem véges számú elemi eseményből áll, akkor analógiát keresünk valamely geometriai alakzat – szakasz, síkidom, test – és az eseménytér között, majd az elemi eseményeket ezek részhalmazaként értelmezzük. Az esemény bekövetkezésének valószínűségét területarányokkal (szakasz- ill. térfogatarányokkal) tudjuk megadni.

Ha egy geometriai alakzattal megadott eseménytérben annak a valószínűsége, hogy egy véletlenül kiválasztott pont az  $A \subset H$  résztartományba esik arányos az  $A$  tartomány mértékével, geometriai valószínűségről beszélünk.

$$P(A) = \frac{A \text{ mértéke}}{H \text{ mértéke}}$$



Tájmetriai vizsgálatokhoz 1 km-szer 1 km-es észak-déli tájolású mintaterületeket határoz meg véletlenszám-generálás segítségével. Mekkora a valószínűsége ennek, hogy egy mintaterület digitális feldolgozásához több EOTR 1:10000-es alaptérképet kell megszereznie, feltételezve a nomál, nem határmenti szelvényeket?

Legyen **A** esemény: a mintaterület egy szelvénylapra esik, **B** esemény: a mintaterület két szelvénylapra esik: **C** esemény: a mintaterület 4 szelvénylapra esik. A három esemény adja a teljes eseményteret.

A vizsgálat akkor korlátozódik egy szelvényre, ha a mintaterület középpontja beleesik a térkép középső,  $5 \cdot 3 \text{ km}^2$ -es részére. Ennek geometriai valószínűsége:

$$p(A) = \frac{5 \cdot 3}{6 \cdot 4} \cdot 100\% = \frac{15}{24} \cdot 100\% = 62,5\%$$

A vizsgálat akkor terjed ki 4 szelvényre, ha a mintaterület középpontja valamelyik sarok  $0,5 \cdot 0,5 \text{ km}^2$ -es részére esik. Ennek geometriai valószínűsége:

$$p(C) = 4 \cdot \frac{0,5 \cdot 0,5}{6 \cdot 4} \cdot 100\% = \frac{1}{24} \cdot 100\% = 4,17\%$$

Tehát annak a valószínűsége, hogy a vizsgált mintaterület egy szelvénylapra esik,  $p(A)=62,50 \%$ , annak, hogy két lapra esik  $p(B)=100 \%-p(A)-p(C)=33,33 \%$ , míg  $p(C)=4,17 \%$ . Ennek segítségével – ismerve egy szelvény árát – kiszámolható az egy mintaterülethez tartozó térképlapok valószínű száma és költsége (feltételezve, hogy egy szelvényre legfeljebb egy mintaterület jut). Az előbbi értéke (d):

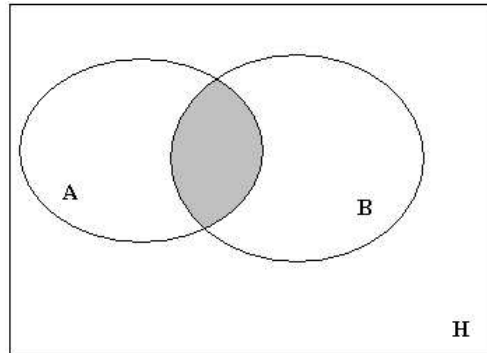
$$d = 0,625 \cdot 1 + 0,3333 \cdot 2 + 0,0417 \cdot 4 = 1,4584$$

Ennek értelmében egy mintaterülethez átlagosan kb. 1,5 térképszelvényt kell megvenni.

## 8. 5. Feltételes valószínűség

*Mekkora a **B** esemény bekövetkezése esetén annak a valószínűsége, hogy egyidejűleg **A** esemény is bekövetkezik?*

Ha azt akarjuk meghatározni, hogy mekkora a  $P(A)$ , ha a **B** esemény bekövetkezett, akkor az eredeti kísérlet körülményei változnak, hiszen a **B** bekövetkezésével a szóba jövő összes elemi események (esetleg a kedvező eseményekét is) összességét leszűkítjük.



Az új esemény:  $A|B$  („A vonás B”)

Az A valószínűsége feltéve, hogy a B bekövetkezik:  $P(A|B)$

Ha  $A, B \subset H$  és  $P(B) \neq \emptyset$ , akkor a  $P(A|B) = \frac{P(A \cap B)}{P(B)}$  hányadost az A eseménynek a B eseményre vonatkoztatott **feltételes valószínűség**ének nevezzük.

I.  $0 \leq P(A|B) \leq 1$

II.  $P(B|B) = 1$

III. Ha  $A_1, A_2, \dots, A_n$  véges sok vagy megszámlálhatóan végtelen sok páronként egymást kizáró események, akkor

$$P(A_1 \cup A_2 \cup \dots \cup A_n | B) = P(A_1 | B) + P(A_2 | B) + \dots + P(A_n | B).$$

### 8. 6. Valószínűségek szorzási szabálya

A és B két tetszőleges esemény, ha  $P(A) > 0$  és  $P(B) > 0$ , akkor az egymásra vonatkoztatott feltételes valószínűségek:

$$P(A|B) = \frac{P(A \cdot B)}{P(B)} \qquad P(B|A) = \frac{P(B \cdot A)}{P(A)}$$

$$P(A \cdot B) = P(A|B) \cdot P(B) \qquad P(B \cdot A) = P(B|A) \cdot P(A)$$

A szorzási szabály általánosan:

$$P(A_1 \cdot A_2 \cdot \dots \cdot A_n) = P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1 \cdot A_2) \cdot \dots \cdot P(A_n | A_1 \cdot A_2 \cdot A_3 \cdot \dots \cdot A_{n-1})$$

## Példa

Egy áruház látogatóinak  $\frac{1}{4}$ -e keresi fel a műszaki cikkek osztályát és ezek közül átlagosan 64% vásárol is. Mi a valószínűsége, hogy egy véletlenül kiválasztott vásárló műszaki cikket vett?

$$A - \text{felkeresi a műszaki osztályt} \quad P(A) = 25\% = 0,25$$

$$B - \text{vásárol ezen az osztályon} \quad P(B) = 64\% = 0,64$$

$$P(A \cap B) = 0,25 \cdot 0,64 = 0,16$$

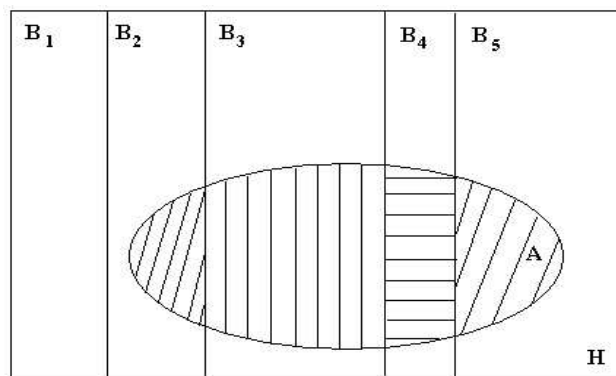
Tehát annak a valószínűsége, hogy egy véletlenszerűen kiválasztott vásárló a műszaki osztályon vásárolt: 16%. Teljesen más a végeredmény, ha így hangzik a feladat: Egy áruház látogatóinak  $\frac{1}{4}$ -e keresi fel a műszaki cikkek osztályát és az áruház látogatói közül átlagosan 64% vásárol is. Mi a valószínűsége, hogy egy véletlenül kiválasztott vásárló műszaki cikket vett? (9%)

## 8. 7. A teljes valószínűség tétele

A valószínűségi mező gyakran felbontható olyan részekre, amelyeket külön-külön már jól tudunk kezelni.

Ha a  $H$  eseménytér  $B_1, B_2, \dots, B_n$  eseményei teljes eseményrendszert alkotnak és  $P(B_k) > 0$  (minden  $k$ -ra), akkor a  $H$ -hoz tartozó bármely esemény valószínűsége:

$$P(A) = \sum_{k=1}^n P(A | B_k) \cdot P(B_k)$$



Tágabb értelemben teljes eseményrendszernek nevezzük az események olyan sorozatát is, amelyek egymást páronként kizárják, és valószínűségeik összege 1.

### Példa

A teljes valószínűség tételét a következő módon alkalmazzuk. A valószínűségi mezőt részekre bontjuk úgy, hogy az egyes részekben belül a (feltételes) valószínűség egyszerűen kiszámítható legyen, és ezeket a valószínűségeket a részek valószínűségeivel súlyozva összeadjuk. Az eljárás pont az, amit különböző koncentrációjú keverékek összeöntésével kapott keverék koncentrációjának kiszámítására használunk.

Három gép gyárt csavarokat. Az első gép 1%, a második 2%, a harmadik 3% selejtet produkál. Az első gép az össztermék 50%-át, a második 30%-át, a harmadik 20%-át állítja elő. Az össztermékből véletlenszerűen választva egyet, mennyi a valószínűsége, hogy az selejt?

A teljes valószínűség tétele alapján a megoldás:

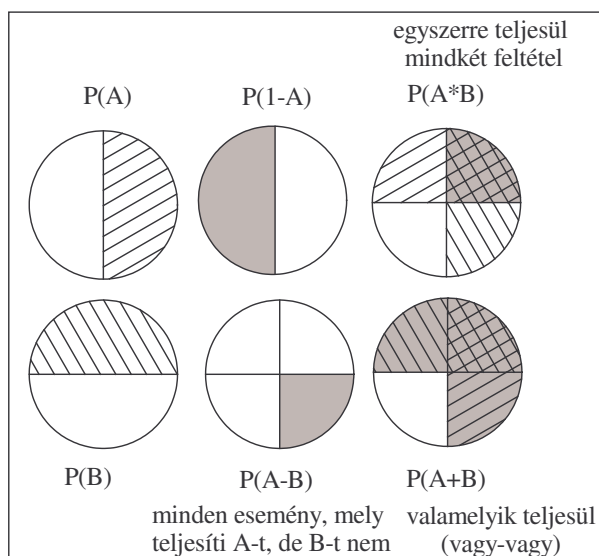
$$P(\text{selejt}) = 0,01 \cdot 0,5 + 0,02 \cdot 0,3 + 0,03 \cdot 0,2 = 0,017 \quad (1,7\%)$$

### 8. 8. Események függetlensége

A és B eseményt akkor tekintjük (sztochasztikusan) függetlennek, ha

$$P(A \cdot B) = P(A) \cdot P(B)$$

Ez azt jelenti, hogy  $P(A|B) = P(A)$ , vagyis B esemény semmilyen befolyással nem bír az A esemény bekövetkezésére.



### Példák:

A valószínűségszámítás egyes elemei levezethetők halmazműveletekből:

**1,** Adva van egy céltábla. Legyen A esemény annak bekövetkezése, hogy a lövés a céltábla jobb felét éri, B esemény pedig reprezentálja azt, hogy a lövés a céltábla felső felét találja el. A két esemény valószínűsége:

$P(A) = 0,5$   $P(B) = 0,5$ , hiszen a céltábla feléről van szó mindkét esetben.

Ha a két esemény egyszerre teljesül, azt A és B szorzatának nevezzük, jelölése AB.  $P(AB)$  tehát annak a valószínűsége, hogy a lövés a céltábla jobb oldalát és a tetejét éri, azaz  $P(A \cdot B) = P(A \cap B) = 0,5 \cdot 0,5 = 0,25$ .

Ha a két esemény közül legalább az egyik teljesül, tehát vagy A vagy B valósul meg, azt A+B-vel jelöljük.  $P(A+B)$  tehát annak a valószínűsége, hogy a lövés a céltábla tetejét vagy jobb felét találja el.  $P(A+B)$  tehát  $P(A \cup B)$ . Ugyanakkor példánkban  $P(A+B)$  nem egyenlő  $P(A)+P(B)$ -vel, hiszen ez  $0,5+0,5=1$  lenne, míg a valóságban a céltáblán  $P(A+B)=0,75$ . Ennek oka az, hogy a két halmaznak van közös része (a metszet, mely a céltábla területének negyedét teszi ki). Helyesen, mivel egymást nem kizáró események összegéről van szó:  $P(A+B) = P(A)+P(B)-P(A \cdot B)$ .

Ha A és B esemény közül csak az egyik megvalósulása lehetséges, (pl. a találat a jobb alsó, vagy a bal felső térrészbe esik), akkor A és B különbségéről beszélünk, tehát A-B és B-A a jelölés.  $P(A-B)$  annak a valószínűségét jelzi, hogy a lövés a céltábla jobb felét éri, de a céltábla alját találja el.

## 2, $P(A|B)$ -t A esemény B-re vonatkoztatott valószínűségének nevezik.

Az együttes bekövetkezés valószínűsége (ha A és B nem függetlenek):

$$P(AB)=P(A|B) \cdot P(B)$$

Együttes bekövetkezés valószínűsége A és B függetlensége esetén:

$$P(AB)=P(A) \cdot P(B)$$

Legyen  $P(A)$  annak a valószínűsége, hogy májusban Salgótarjánban csapadékos nap következik be,  $P(B)$  pedig annak a valószínűsége, hogy májusban Budapesten következik be esős nap.  $P(A)=0,42$ ,  $P(B)=0,40$  a mérések alapján.  $P(A|B)=0,71$ , tehát 71% a valószínűsége, hogy amikor Salgótarjánban esik, akkor Budapesten is. Belátható, hogy a két esemény bekövetkezése nem független egymástól, hiszen  $P(A|B) > P(A)$ . a két esemény együttes előfordulása:  $P(AB)=P(A|B) \cdot P(B)=0,71 \cdot 0,4=0,29$ . Abban az esetben, ha Salgótarjánban és Budapesten egymástól függetlenül esne az eső, akkor  $P(AB)=P(A) \cdot P(B)=0,17$ , tehát kisebb az együttes bekövetkezés valószínűsége.

## 3, Egymást nem kizáró események összege:

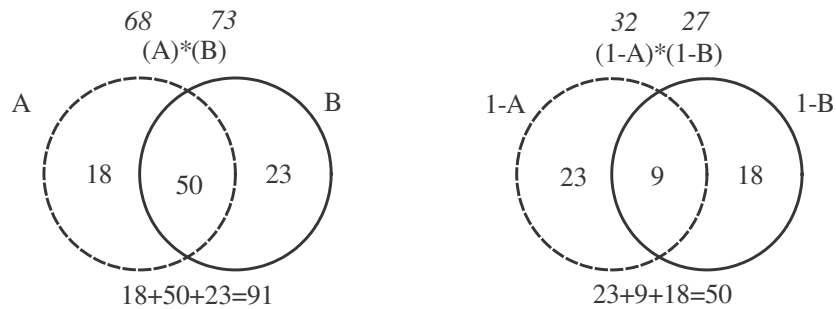
$$P(A+B)=P(A)+P(B)-P(AB)$$

Ellentétes esemény valószínűsége:

$$P(A')=1-P(A)$$

Legyen A esemény az, hogy Budapesten július 1-én nem esik az eső, B esemény pedig, hogy július 1-én Szegeden nem esik az eső.  $P(A)=0,68$ ,  $P(B)=0,73$ . Annak a valószínűsége, hogy legalább az egyik helyen nem esik:  $P(A+B)=P(A)+P(B)-P(AB)=0,68+0,73-0,5=0,91$ , azaz 91%. Ugyanígy annak

valószínűsége, hogy legalább egyik helyen esik az eső július 1-én:  
 $P(1-A+1-B) = P(1-A) + P(1-B) - P((1-A) \cdot (1-B)) = 0,32 + 0,27 - 0,09 = 0,5$ , azaz 50%. A fenti feladat halmazokkal történő megközelítését a következő két ábra mutatja meg (összesen 100 eseményt feltételezve):



- 4, Adott az eseménytérben (Kopárosvölgy falu 14 háza) 2 elemi eseménysor (v. halmaz): A a munkanélküli férjeket, B a munkanélküli feleségeket reprezentálja (páronként összetartozó adatok). Ekkor annak a valószínűsége, hogy munkanélküli férfival / nővel találkozunk, ha véletlenszerűen bekopogtatunk egy házba:

$$P(A) = 7/14 = 0,5 \quad P(B) = 6/14 = 0,42.$$

Annak az esélye, hogy olyan házba kopogtatunk be, ahol a férj és feleség is munkanélküli:  $P(A \cdot B) = 4/14 = 0,28$  (az együttes teljesülések kiválogatva az adatsorból).

Ha A és B függetlenek lennének, akkor  $P(A \cdot B) = P(A) \cdot P(B) = 42/196 = 0,21$ , (itt nem független egymástól a két elemi esemény, mint ez a feladatból kiderül).

A táblázatból leolvasható, hogy a 6 munkanélküli feleséghez 4 munkanélküli férj tartozik, azaz annak a valószínűsége, hogy a férj munkanélküli, ha a feleség állástalan  $P(A | B) = 66\%$ .

$P(A | B) = 4/6 = 0,66$        $P(B | A) = 4/7 = 0,56$  amit megkaphatunk úgy is, hogy  $P(A | B) = P(A \cdot B) / P(B) = 0,28 / 0,42 = 0,66$

$$P(B | A) = P(A \cdot B) / P(A) = 0,28 / 0,5 = 0,56$$

Ha helytelenül a 0,21-es értékkel számolnánk P(A)-hoz és P(B)-hez jutnánk vissza:  $0,21 / 0,42 = 0,5$ .

A	B
n	i
n	n
i	i
i	n
n	i
n	n
i	n
n	n
i	i
n	n
n	n
i	i
i	n
i	i

- 5, Teljes eseményrendszer elemi eseményeinek összege:

$$P(A_1) + P(A_2) + P(A_3) + \dots = 1$$

Azaz az egyik bekövetkezte biztos esemény, egymást viszont kizárják a lehetőségek. Legyen  $P(A_1)$  a párasság bekövetkeztének valószínűsége 0,2,  $P(A_2)$ , a köd bekövetkeztének valószínűsége 0,4, ekkor tehát a jó látáskörülmények esélye (ha nincs több befolyásoló tényező):

$$P(A_3) = 1 - (P(A_1) + P(A_2))$$

### **Feladatok**

- 1, Egy békési mintaterületen a kutak 30%-a arzénnal erősen szennyezett, 60%-a mérsékelten szennyezett, 10%-a tiszta minősítést kapott. Az arzénnal erősen szennyezett vízből ivók 30%-a, a mérsékelten szennyezőkből ivók 20%-a és a tiszta vízből ivók 10%-a kapott mérgezést. Mekkora esélyem van mérgezést kapni, ha nem tudom, hogy milyen besorolású kútból iszom?
- 2, Talajvizsgálathoz kiválasztott mintaterület (egy teljes 1:10000-es EOTR-szelvény, 24 km<sup>2</sup>) felszínborítottságát a következő értékek jellemzik: 33,5 % erdő, 25 % rét, 40 % szántó, 1,5 % zártkert, 10 % felszíni víz. A zártkertek egy négyzet alakú terület egységbe tömörülnek, a mintaterület középső részén. A felszíni víz két egyforma méretű, kör alakú kaldera-tó. Mindkét területtípus alkalmatlan mintavételezésre. Előzetes vizsgálatainkhoz az 1·1 km<sup>2</sup>-es egységek közepéről gyűjt mintát. A minták hány százalékát tudja begyűjteni? A minták helyei közül találmra kiválasztva 8-at mi annak a valószínűsége, hogy valamelyik minta nem gyűjthető be?  
Ezen a területen más pontméréseket is végezne, melyekből ki kell zárni a szegélyzónákat (mely a foltok közötti határ mindkét oldalán átlagosan 5 m-nek határoz meg). A legegyszerűbb geometriát feltételezve mennyi a minimális esélye annak, hogy a véletlenszerűen kiválasztott pont a mérésre alkalmatlan?
- 3, 1:20000-es méretarányú digitális tematikus térképén a közutakat egy 1 mm széles jellel ábrázolja. 10 km<sup>2</sup>-es mintaterületén a közúthálózat-sűrűség 1,8 1/km, míg az útkereszteződések száma 2 1/km<sup>2</sup>. Tetszőlegesen kiválasztott pontmintánk koordinátáit a megfelelő fólia adatbázisába rögzítve megjelenítjük térképünkön. Mi annak az esélye, hogy az 1 mm pontnagyságú jel takarja az utak valamely részét?
- 4, Alföldi területet ábrázoló (1:10000-es, 6 km<sup>2</sup>-es) digitális domborzati térképén minden főszintvonalra (12 db) felírja a szintvonal értékét, egy címke a térképen 4·8 mm<sup>2</sup>-nyi helyet foglal el. Mi annak a valószínűsége, hogy a címkék kitarják a véletlenszerűen elhelyezkedő besugárásmérő állomásaink (8 db) valamelyikét?
- 5, Egy ingázók szokásait felmérő kérdőívezéshez keres alanyokat egy nagy gépgyártó üzemben. Az „A” településről 15 római katolikus és 15 református, a „B” településről 15 római katolikus és 5 református, „C” településről 30 római katolikus és 10 református vallású dolgozó jár be az üzembe naponta. Az üzemcsarnokban jelen van az összes ingázó. Mi a valószínűsége annak, hogy egy találmra kiválasztott ingázó római katolikus? Mi a valószínűsége annak, hogy egy találmra kiválasztott ingázó a B településről jár be és református?





## 9. Folytonos valószínűségi változók eloszlásai

A **valószínűségi változó** a valószínűségszámítás egyik legfontosabb fogalma. Lényegében olyan jelenségek matematikai megfogalmazására, modellezésére alkalmas, melyek véletlentől függő értéket vesznek fel. Ilyen lehet például egy kockadobás eredménye, egy folyó vízállása, vagy az utcán szembe jövő emberek testmagassága.

A **valószínűségi változó lehet:**

**a, diszkrét** - olyan, ami elkülönült értékeket tud csak felvenni (pl. egy autóbusszra felszálló 44 utasból mennyi a férfi)

**b, folytonos:** olyan, ami egy intervallumon belül- bármilyen értéket felvehet (pl. a következő felszálló utas milyen magas).

Folytonos valószínűségi változónál az eloszlást az ún. **sűrűségfüggvény** adja meg. Ennek a formulája alapján a legkülönbözőbb eloszlások lehetségesek. Ezek közül tekintünk át néhányat.

### 9. 1. Egyenletes eloszlás

Ha egy valószínűségi változó az  $[a, b]$  intervallum valamennyi értékét azonos valószínűséggel veszi fel, akkor egyenletes eloszlású:

Sűrűségfüggvénye:

$$f(x) = \begin{cases} 0 & \text{ha } x < a \\ \frac{1}{b-a} & \text{ha } a \leq x \leq b \\ 0 & \text{ha } x > b \end{cases}$$

Eloszlásfüggvénye:

$$F(x) = \begin{cases} 0 & \text{ha } x < a \\ \frac{x-a}{b-a} & \text{ha } a \leq x \leq b \\ 1 & \text{ha } x > b \end{cases}$$

$$\text{Várható értéke: } M(\xi) = \frac{a+b}{2}$$

$$\text{Szórása: } D(\xi) = \frac{b-a}{2\sqrt{3}}$$

**Példa:**

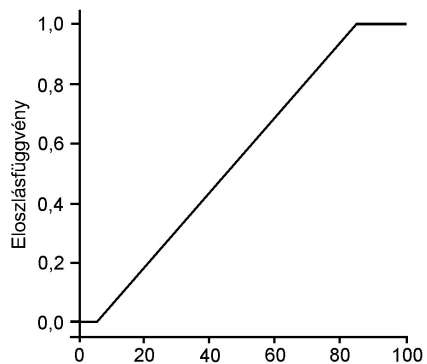
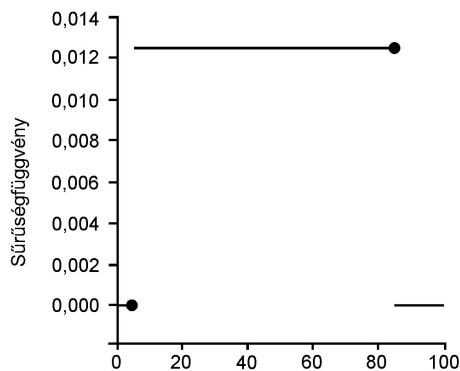
Egy nyírségi homokos rétegsor mintáinak áteresztőképesség-vizsgálatánál azt tapasztalják, hogy az első vízcsepp megjelenéséig a mérőhenger alján az eltelt idő 5 és 85 másodperc között van. Az eltelt idő legyen az **egyenletes eloszlású valószínűségi változó**. Határozzuk meg a valószínűségi változó sűrűség- és eloszlásfüggvényét.

Sűrűségfüggvény

$$f(x) = \begin{cases} 0, & \text{ha } x \leq 5 \\ \frac{1}{80} & \text{ha } 5 < x \leq 85 \\ 0 & \text{ha } x > 85 \end{cases}$$

Eloszlásfüggvény

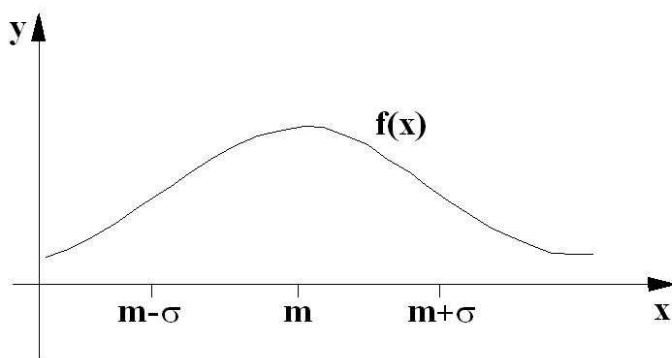
$$F(x) = P(X < x) = \begin{cases} 0, & \text{ha } x \leq 5 \\ \frac{x-85}{80} & \text{ha } 5 < x \leq 85 \\ 1 & \text{ha } x > 85 \end{cases}$$



szívargási idő (s)

## 9. 2. Normális eloszlás

A valószínűségi változó eloszlása normális, ha sűrűségfüggvénye szimmetrikus haranggörbe, amely az  $m$  és  $\sigma$  paraméterekkel jellemezhető.



$$\text{Formulája: } f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

Maximuma az  $x = m$  helyen van. A  $\sigma$  érték a lapultságra (szórás/félértékszélesség) jellemző szám.

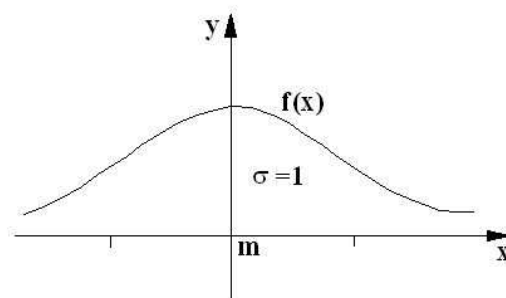
Eloszlásfüggvénye:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-m)^2}{2\sigma^2}} dt$$

Várható értéke:  $M(\xi)=m$  Szórása:  $D(\xi)=\sigma$

Mivel az eloszlásfüggvény nehezen számolható ki, ezért ezt táblázatból, vagy számítógéppel szokták meghatározni. Az eloszlásfüggvény visszavezethető a standard normális eloszlásra. (pl. Mérési hiba eloszlása, egy „gyártósoron” készült alkatrészek méreteloszlása, azonos korú gyerekek magasságeloszlása általában normális eloszlású.)

**Standard normális eloszlásúnak** nevezzük a normális eloszlású valószínűségi változót, ha  $m=0$  és  $\sigma=1$



Ha egy valószínűségi változó értékét nagyszámú, egymástól függetlenül ható véletlen tényező határozza meg úgy, hogy az egyes tényezők külön-külön csak igen kis mértékben járulnak hozzá az összes véletlen hatásból eredő ingadozáshoz, és az egyes tényezők hatásai összeadódnak, akkor általában normális eloszlású valószínűségi változót kapunk. Pl. skálán mérhető jellemzők (hőmérséklet, stb.)

**Példa:**

- 1, Szombathelyen az áprilisi középhőmérséklet  $9,7\text{ }^\circ\text{C}$ , a középhőmérsékletek szórása  $2,3\text{ }^\circ\text{C}$ . Mennyi a valószínűsége annak, hogy az áprilisi középhőmérséklet  $8\text{ }^\circ\text{C}$ -nál alacsonyabb?

Mivel a középhőmérsékletek eloszlása normális, ezért *normális eloszlással kapcsolatos feladatról* van szó.

$$m = 9,7\text{ }^\circ\text{C}$$

$$\sigma = 2,3\text{ }^\circ\text{C}$$

a, Oldjuk meg *grafikusan* a feladatot!

Fölrajzoljuk az  $m$  és  $\sigma$  paraméterekkel megadott normális eloszlás sűrűségfüggvényét. Bejelöljük a vízszintes tengelyen a  $8^\circ\text{C}$ -ot, s az annál kisebb értékekkel jellemzett görbe alatti terület a keresett valószínűség, azaz  $F(x_a = 8^\circ\text{C})$ .

Figyelem! A görbe alatti összeterület balról jobbra kumulálva =1 vagy 100%, ennek azon részéről van tehát szó, amelyet a  $8^\circ\text{C}$ -os vonal a teljes területből kimetsz.

b, *Numerikus megoldás:*

Ehhez az eredeti képleten transzformációt kell végrehajtani:

$$\frac{x_a - m}{\sigma} = d_a$$

Ez a transzformáció tetszőleges  $n(m \neq 0; \sigma \neq 1)$  eloszlást  $n(m = 0; \sigma = 1)$  eloszlássá (**standard normális eloszlássá**) alakít. Erre a transzformációra azért van szükség, mert az  $n(m = 0; \sigma = 1)$  eloszlásnak, és csak ennek ismert az eloszlásfüggvénye. Azaz az  $F(x_a) = P(x < x_a)$  minden egyes  $x_a$ -ra csak az  $n(m = 0; \sigma = 1)$  eloszlásra ismert, mely táblázatosan adott. Ily módon bármelyik normális eloszlással kapcsolatos feladat ennek a transzformációnak a segítségével oldható csak meg. Ez a transzformáció egy kölcsönösen egyértelmű megfeleltetés, mely az eredeti eloszlás minden egyes értékéből egy új értéket képez. Tehát:

$$(8 - 9,7) / 2,3 = -0,73$$

Számunkra az  $x_a = 8^\circ\text{C}$ , illetve az ennek megfelelő transzformált érték a transzformációval előállított  $d_a$  független változóhoz tartozó  $F(d_a)$  függvényérték a standard normális eloszlás eloszlásfüggvényének a táblázatából kikereshető (**melléklet:** a standard normális eloszlású valószínűségi változó eloszlásfüggvényének táblázata). Negatív értékek nincsenek e táblázatban, mivel azonban  $\Phi(-x) = 1 - \Phi(x)$ , így a  $-0,73$  helyett megnézzük a  $0,73$ -hoz tartozó függvényértéket  $(0,76)$ , majd ezt kivonjuk 1-ből. Így  $p=0,23$ , azaz 23%-os a valószínűsége, hogy a hőmérséklet a megadott értéknél alacsonyabb.

2, Egy harmadik világbeli országban a bányászatban dolgozók napi munkaidejét normális eloszlásúnak találták  $m = 14$  átlaggal és  $\sigma = 2$  szórással. Egy ilyen populációban mi annak a valószínűsége, hogy a napi munkaidő 10 és 15 óra közé esik? (A dolgozók hány %-a dolgozik 10–15 órát?)

*Megoldás:* standardizálás után a  $m = 14$ -nek megfelel a  $z = 0$ , A 2 standard deviációnak (szórásnak) pedig az 1. Alkalmazzuk a  $z$  transzformációt a 15-re és a 10-re, kapjuk a következő standardizált értékeket:

$$x_{15}=(15-14)/2=0.5 \text{ és } x_{10}=(10-14)/2 = -4/2= -2.$$

$F(0.5)=0.6915$  és  $F(-2)=0.0228$ . Kivonás után (mivel a két érték közötti területről van szó, a két érték pedig a valószínűség kumulált értéke, azaz a görbe alatti terület 0-tól x-ig)

$$p= 0.6915-0.0228=0.6687.$$

Tehát várhatóan a bányászok 67 %-ának fog 10 és 15 óra közé esni a munkaideje.

### 9. 3. Lognormális eloszlás

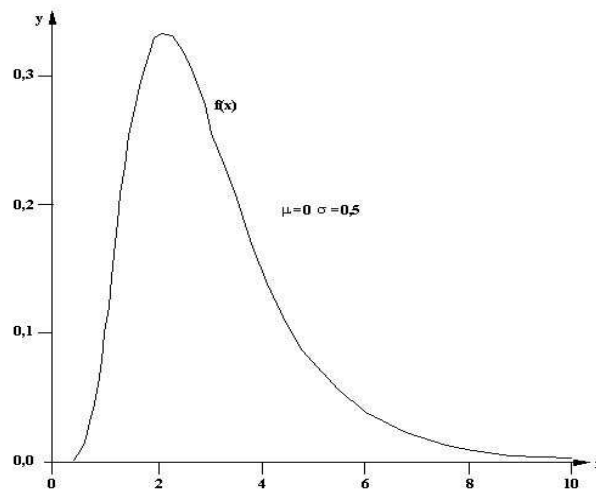
Egy valószínűségi változó lognormális eloszlású, ha a változó logaritmus a normális eloszlású.

$$\varphi = \ln \xi \text{ és sűrűségfüggvénye } f(x) = \frac{1}{\sqrt{2\pi\sigma x}} \cdot e^{-\frac{(\ln x - m)^2}{2\sigma^2}}, \quad x > 0$$

Az eloszlás várható értéke és varianciája:

$$M(\xi) = \mu = e^{m + \frac{\sigma^2}{2}} \quad D^2(\xi) = \sigma^2 = e^{2m + \sigma^2} (e^{\sigma^2} - 1)$$

A statisztikában gyakran alkalmazott eljárás, ha a minta eloszlása nem normális, akkor vesszük a mintaelemek logaritmusát. Ezzel a művelettel gyakran sikerül az adatokat normális eloszlásúvá transzformálni.



Lognormális eloszlású görbe

Ha egy feladatban a valószínűségi változó csak pozitív értékeket vehet fel és a várható érték nem túl magas (közel van a 0-hoz), és ehhez képest a szórás nagy, akkor az eloszlás aszimmetrikussá válik. Ez úgy mutatkozik meg, hogy a

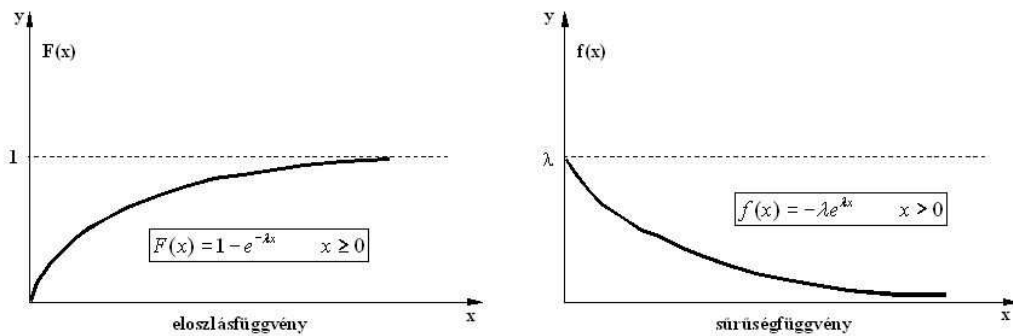
görbe a pozitív irányba eltorzul, elnyúlik. Ha a variációs koefficiens 0,5-nél kisebb, akkor általában normális eloszlást kapunk. Ha a variációs koefficiens 1 vagy annál nagyobb, a csak pozitív értékeket felvehető valószínűségi változók eloszlása torzult lesz, és nagyon sok esetben lognormális eloszlással lesz megközelíthető:  $y = \lg(x+1)$ .

#### 9. 4. Exponenciális eloszlás

Egy valószínűségi változó exponenciális eloszlású, ha eloszlásfüggvénye:  $F(x) = 0$ , ha  $x = 0$  és  $F(x) = 1 - e^{-\lambda x}$  ha  $x \geq 0$ ; sűrűségfüggvénye:  $f(x) = \lambda$ , ha  $x = 0$ ;  $f(x) = \lambda e^{-\lambda x}$ , ha  $x > 0$  és  $f(x) = 0$ , ha  $x = \infty$ . A sűrűségfüggvény és az eloszlásfüggvény kapcsolata alapján az exponenciális eloszlású valószínűségi változó eloszlásfüggvénye:

$$F(x) = F(\xi < x) = 1 - e^{-\lambda x}$$

Exponenciális eloszlást követnek a különféle várakozási idők, például a radioaktív bomlás során az egyes atomok élettartama. Ugyancsak exponenciális eloszlásúak a használati tárgyak vagy azok különböző alkatrészeinek élettartamai.



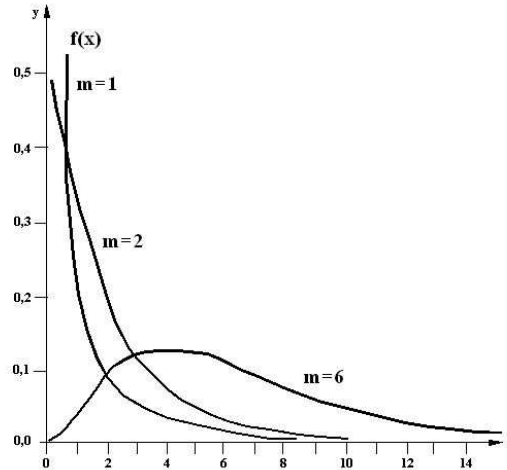
Az exponenciális eloszlású valószínűségi változó várható értéke és varianciáját a következő összefüggések adják meg:

$$M(\xi) = \mu = \frac{1}{\lambda}$$

$$D^2(\xi) = \sigma^2 = \frac{1}{\lambda^2}$$

## 9. 5. $n$ szabadságfokú $\chi^2$ -eloszlás

Legyenek  $x_1, x_2, \dots, x_n$  független standard normális eloszlású valószínűségi változók. Ekkor  $y = x_1^2 + \dots + x_n^2$  eloszlását  $n$  szabadságfokú  $\chi^2$ -eloszlásnak nevezzük.



$n$  számú független  $N(0, 1)$  eloszlású valószínűségi változó négyzetgyökének eloszlását  $n$  szabadságfokú  $\chi^2$  eloszlásnak nevezzük.  $n$  értékétől függően a sűrűségfüggvény különböző alakú lehet.

## 9. 6. Student- vagy t-eloszlás

Ha van egy olyan valószínűségi változónk, amely egy standard normális eloszlású valószínűségi változó  $\sqrt{n_t}$  - szerese, valamint egy másik, ettől független olyan valószínűségi változó, amely egy  $n_t$  szabadságfokú  $\chi^2$ -eloszlású valószínűségi változó négyzetgyöke, akkor a kettő hányadosát  $n_t$  szabadságfokú Student- vagy t-eloszlásúnak nevezzük.

$$t = \frac{\eta \sqrt{n_t}}{\sqrt{\xi_1^2 + \xi_2^2 + \dots + \xi_{n_t}^2}}$$

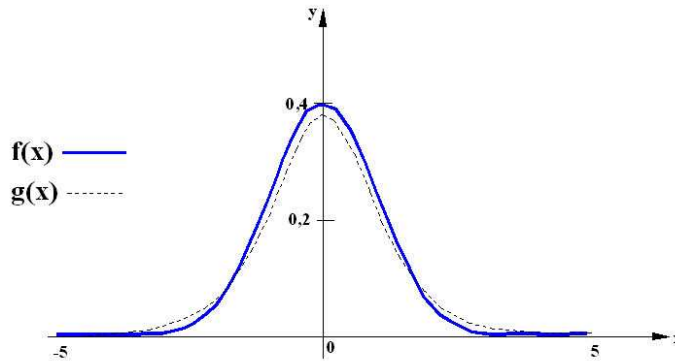
A Student eloszlást a geostatisztikában nem elsődlegesen a paraméterek eloszlásának közelítésére, hanem például szignifikancia vizsgálatoknál használjuk.

A Student eloszlás sűrűségfüggvénye  $\forall x \in \mathbb{R}$  esetén  $f_n(x)$ , ahol  $\Gamma$  a teljes gamma függvény. Ha  $n_t > 80$ , a normális eloszlás jól közelíti a Student eloszlást.

$$f_n = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \cdot \frac{1}{\left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}}$$

A Student-eloszlás 0 pontra nézve egy szimmetrikus eloszlás. Az  $n_t$  szabadságfokú Student eloszlású valószínűségi változó várható értéke:  $n_t = 1$  esetén nem létezik,  $n_t \geq 2$  esetén  $M(t) = 0$ , szórásnégyzete pedig:  $n_t = 1$  és  $n_t = 2$  esetén nem létezik,  $n_t \geq 3$ -nál  $D^2(t) = n/(n-2)$ .

A Student vagy  $t$ -eloszlás a várható értékre szimmetrikus.



Az  $N(0,1)$  eloszlás ( $f(x)$ ) és az  $n=5$  szabadsági fokú Student-eloszlás ( $g(x)$ ) sűrűségfüggvénye

## 9. 7. Binomiális eloszlás

A megfigyelésünk kétféle eredményt adhat: egy  $A$  esemény bekövetkezik, vagy nem következik be.  $n$  számú (egymástól független) megfigyelést végzünk. Az  $A$  esemény valószínűsége:  $p$ . A valószínűségi változó:  $n$  megfigyelésből az  $A$  esemény  $k$  szor következik be ( $\xi=k$ ).

$$P_k = \binom{n}{k} \cdot \frac{S^k (N-S)^{(n-k)}}{N^n}$$

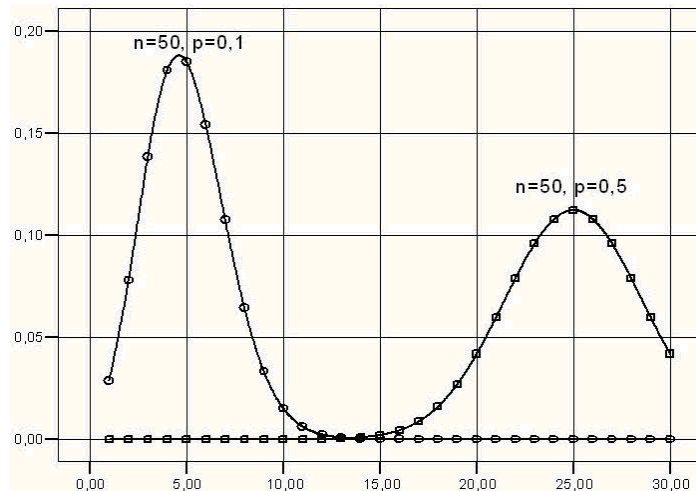
Ez megfelel az alábbiaknak:

Annak a valószínűsége, hogy az  $A$  esemény  $n$  számú megfigyelésből  $k$ -szor következik be:

$$P(\xi = k) = \binom{n}{k} \cdot p^k (1-p)^{n-k}$$

A binomiális eloszlásban az  $n$  és a  $p$  un. paraméterek.





**Példa:**

- 1, Magyarországon szeptemberben az anticiklon-centrum **makroszinoptikus** helyzet bekövetkezési valószínűsége 21 %. Egymást követő 4 napot kiválasztva adjuk meg annak a valószínűségét, hogy ebből a 4 napból 3 napon fordul elő anticiklon-centrum időjárási helyzet!

*Alapkérdés:* Adott egy  $p$  alapvalószínűségű alternatív esemény.  $n$  esetből  $k$ -szori bekövetkezése milyen valószínűséggel várható? Mivel az alapkérdés alkalmazható a feladatra, ezért az vagy a binomiális-, vagy a Poisson-eloszlás segítségével oldható meg.

Döntsük el a feladat típusát! Ha  $p > 0,03$  akkor binomiális eloszlással számolunk. Ha  $p < 0,03$  akkor Poisson-eloszlással számolunk. Mivel  $p = 21 \% = 0,21$  ezért az *eloszlás típusa: binomiális*. Határozzuk meg a binomiális eloszlás paramétereit!  $n = 4, k = 3$

Helyettesítsünk be a binomiális eloszlás képletébe!

$$P(k, n) = \frac{4!}{3!(4-3)!} \cdot 0,21^3 (1-0,21)^{4-3} \qquad P(k, n) = \frac{n!}{k!(n-k)!} \cdot p^k (1-p)^{n-k}$$

$$P(3; 4) = 4 \cdot 0,21^3 \cdot 0,79 = 0,029$$

Tehát annak a valószínűsége, hogy Magyarországon a kiindulási feltételek mellett egymást követő 4 nap közül 3 napon fordul elő anticiklon-centrum időjárási helyzet:  $p = 0,029 = 2,9\%$ .

**9. 8. Hipergeometrikus eloszlás**

A visszatevés nélküli mintavételnél  $n$  elem között  $S$  rendelkezik egy adott tulajdonsággal. Az adott tulajdonságú elem kiválasztásának valószínűsége:

$$p = \frac{S}{N}$$

$n$  elemből választottunk ki  $n$  számú elemet. A valószínűségi változó:  $n$ -ből  $k$  db. felel meg valamilyen tulajdonságnak:  $\xi = k$

Annak a valószínűsége, hogy a kiválasztott  $n$  elemből pontosan  $k$  számú elem rendelkezik az adott tulajdonsággal:

$$P(\xi = k) = \frac{\binom{S}{k} \cdot \binom{N-S}{n-k}}{\binom{N}{n}}$$

A hipergeometrikus eloszlás paraméterei:  $N, S, n$

Várható értéke: 
$$M(\xi) = \sum_{k=0}^n k \cdot \frac{\binom{n}{k} \cdot \binom{N-S}{n-k}}{\binom{N}{n}} = n \cdot p \quad \left(p = \frac{S}{N}\right)$$

Szórása: 
$$D(\xi) = \sqrt{np(1-p) \frac{N-n}{N-1}}$$

## 9.9. Poisson eloszlás

A ritka események valószínűségi eloszlása. Tekinthető a binomiális eloszlás speciális határértékének, amikor is  $n$  (a megfigyelések száma) nagyon nagy és  $p = P(A)$  – nagyon kicsi.

Akkor a  $P(\xi) = \binom{n}{k} p^k q^{n-k}$  kifejezés jól közelíthető annak határértékével,

ha  $n$  eléggé nagy és  $p$  viszonylag kicsi. Ha  $q = (1-p)$  és az  $n \cdot p = \lambda$ , akkor

$$\lim_{n \rightarrow \infty} \binom{n}{k} p^k (1-p)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}$$

A Poisson eloszlás eloszlásfüggvénye:

$$P(\xi) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$$

(Ezzel sokszor könnyebb számolni, mint a binomiális eloszlás képletével.)

Paramétere:  $\lambda$  ahol:  $\lambda = n \cdot p$

Várható értéke:  $M(\xi) = \lambda$  Szórása:  $D(\xi) = \sqrt{\lambda}$

(pl. Mikroszkóp alatt adott  $\text{mm}^2$ -en leszámolható pollenek száma - ábra)



Véletlen kétdimenziós ponteloszlás (Poisson-eloszlás: pl. emberek a téren, fák egy természetes erdőben, stb.)

### Példák

- 1, Intenzív zivatarok esetén átlagosan 10 másodpercenként észlelhető egy villámlás. Poisson-féle eloszlást feltételezve mi annak a valószínűsége, hogy 15 másodperc alatt kettő villámlást észlelünk?

*Megoldás*

Az adatok alapján  $p = 0,1$ ,  $n = 15$ , valamint  $\lambda = 1,5$  és  $k = 2$ . A keresett valószínűség:

$$P = \frac{1.5^2 e^{-1.5}}{2!} = 0.25$$

- 2, Intenzív zivatarok esetén átlagosan 10 másodpercenként észlelhető egy villámlás. Poisson eloszlást feltételezve mi annak a valószínűsége, hogy 15 másodperc alatt kettőnél több villámlást észlelünk?

*Megoldás*

A kívánt esemény akkor következik be, ha az észlelt villámlások száma vagy 3, vagy 4 stb. Könnyű belátni, hogy a villámlás 0-szor, 1-szer, 2-szer stb. történő bekövetkezése teljes esemény rendszer (valamelyik biztosan bekövetkezik, és az egyik kizárja a másikat). Így felírhatjuk, hogy:

$$P(0) + P(1) + P(2) + P(3) + \dots + P(n) \dots = 1$$

Azaz a keresett valószínűség:

$$P(3) + P(4) + P(5) + \dots + P(n) \dots = 1 - P(0) + P(1) + P(2)$$

A feladatot könnyebben meg tudjuk oldani, ha  $k = 0, 1$  és  $2$  értékhez tartozó valószínűségeket kiszámoljuk. Az előző feladat alapján  $P(k=0) = 0,223$  (fontos tudni, hogy  $0! \equiv 1$ ),  $P(k=1) = 0,335$  és  $P(k=2) = 0,25$ .

A keresett valószínűség:  $1 - 0,223 - 0,335 - 0,25 = 0,192$ .

- 3, Egy országban az elmúlt száz évben 500 alkalommal történt hármás ikerszülés. Mi a valószínűsége annak, hogy egy évben ez az esemény 2-szer fordul elő.

*Megoldás*

**Az eloszlás  $\lambda$  paraméterének értéke 5, a  $k = 2$ . Így a megoldás:**

$$P = \frac{5^2 e^{-5}}{2!} = 0,08$$

- 4, Nyíregyházán áprilisban 50 év alatt összesen 28 napon fordult elő fagy. Mi a valószínűsége a fagymentes áprilisknak Nyíregyházán?

*Megoldási menet:*

Alapkérdés: Adott egy  $p$  alapvalószínűségű alternatív esemény.  $n$  esetből  $k$ -szori bekövetkezése milyen valószínűséggel várható? Mivel az alapkérdés alkalmazható a feladatra, ezért az vagy a binomiális, vagy a Poisson-eloszlás segítségével oldható meg.

Döntsük el a feladat típusát!

Ha  $p > 0,03$  akkor binomiális eloszlással számolunk

Ha  $p < 0,03$  akkor Poisson-eloszlással számolunk

**Mivel a keresett esemény, azaz a fagyos áprilisi napok száma:  $k = 28$  nap, az összes esemény, azaz az összes vizsgált áprilisi napok száma:  $n = 50$  év  $\cdot$  30 nap = 1500 nap.  $\lambda = np = 0,018 \cdot 30 = 0,56$**

**Ezért:**

$$p = \frac{k}{n} = \frac{28}{50 \cdot 30} < \frac{30}{50 \cdot 30} = \frac{1}{50} = \frac{2}{100} \quad 0,018 < 0,03$$

**Határozzuk meg a Poisson eloszlás paramétereit!  $k = 0$  (azaz 0 napon következzen be fagy egy tetszőleges –  $n = 30$  nap tartamú – áprilisban)!**

$$P(0,0,56) = \frac{0,56^0 \cdot e^{-0,56}}{0!} = e^{-0,56} = \frac{1}{e^{0,56}} = \frac{1}{2,72^{0,56}} = 0,571$$

Tehát annak a valószínűsége, hogy Nyíregyházán áprilisban a kiindulási feltételek mellett ne forduljon elő fagyos nap:  $p = 0,571 = 57\%$

### ***Feladatok***

#### ***Binomiális eloszlás***

- 1, A Tisza partjának gátjait az elmúlt 20 év alatt 6-szor mosta el az árvíz. Mi a valószínűsége, hogy a következő 10 éven belül ez a, pontosan kétszer

b, legalább kétszer megtörténik?

*(Figyelem! Itt a helyes megoldást a  $2+3+4+5+6+7+8+9+10$  esetek valószínűségeinek összege adja. Mivel ezt kiszámolni időigényes, egyszerűbb kiszámolni a  $0+1$  esetek valószínűségét, majd ezt kivonni 100%-ból. Ebben az esetben ugyanazt az értéket kapjuk, mintha a hosszabb számítást végeztük volna el!)*

- 2, A DVSC az elmúlt 10 mérkőzésből 8-at megnyert. A bajnokságból 5 forduló van hátra, a DVSC-nek 3 győztes mérkőzés is elég a bajnoki címhez. Mekkora az esélye annak, hogy a Loki bajnok lesz (kizárva a többi riválist és a döntetleneket a vizsgálatból)?

*(Figyelem! A Loki akkor is bajnok, ha 3, 4 vagy 5 mérkőzést nyer!)*

- 3, A Kесе-patak az elmúlt 10 évben 3-szor mutatott romló vízminőséget az előző évhez képest. Mi az esély arra, hogy a következő 5 évben ez legalább kétszer megtörténik?
- 4, A Balaton partján 100 éves adatsor alapján júniusban a csapadékos napok átlagos száma 5. Mi a valószínűsége annak, hogy egy hetes nyaralásunk alatt 3-szor fog esni az eső?
- 5, Egy folyó vízszintje 50 évente 3-szor haladja meg a 10 m-es szintet. Mi a valószínűsége annak, hogy ez az elkövetkező 10 évben 4-szer fog előfordulni? Mi a valószínűsége annak, hogy ez az elkövetkező 10 évben 2-nél többször fog előfordulni?
- 6, A Bükkben az őserdőben 100 fa közül 3 beteg. Mi az esély arra, hogy 50 fa között pontosan 2 beteg fa van? Mekkora a valószínűségek között a különbség, ha binomiális és ha Poisson eloszlást feltételezünk?
- 7, 100 hektáron a lineáris erózió átlagosan 4 hektárt érint. Parcellánk 20 hektáros, s művelése értelmetlen, ha az erózió 10 hektárt v. annál többet érint. Mi az esélye annak, hogy 20 hektáros földünkön a lineáris erózió 10 hektárnál kevesebbet érint?

### **Poisson-eloszlás**

- 1, Egy környezetkutató cég mintavételei közül átlagosan 100-ból 2 hibás. Mi az esélye annak, hogy egy 10 mintavételes kutatás során 1 hibát követnek el? (Ekkor ugyanis újra el kell végezni a kutatást, s az többletköltséget jelent).
- 2, Lövészárok harc során pergőtűzben átlag 60 másodpercenként hal meg egy katona. Mi az esélye annak, hogy egy 40 fős szakasz megsemmisül egy óra leforgása alatt, mielőtt az erősítés befutna?

*(Figyelem! Egy óra 60·60 másodperc azaz 3600! Használjuk a Poisson-eloszlást,  $p < 0,03$ . Itt a „kedvező” esetek száma, a  $k$ , a halottak száma! Azaz számoljuk ki, hogy mi az esélye annak, hogy mind a 40 katona meghal! )*

- 3, Minden századik 10 tonnánál nagyobb tömeget megmozgató csuszamlásnak volt legalább 1 halálos áldozata. Mekkora esély van arra, hogy a következő 10 ilyen csuszamlásnak legalább 2 halálos áldozata lesz?
- 4, A Földön minden százezer négyzetkilométerre jut egy 100 000 fős nagyváros. Poisson eloszlást feltételezve, mi az esély arra, hogy fele ekkora területen 2 ekkora város van? Mi az esély arra, hogy 100 000 négyzetkilométeren egyetlen ilyen város sincs?

### *Normáleloszlás*

- 1, Az Aranykalász kft. földjének termése az elmúlt években a következő volt: 6, 5, 6, 7, 9, 6, 5, 7, 8, 11, 5, 6 és 10 mázsa kukorica. 7 mázsa fölötti termés esetén a tsz. haszonnal zárja az évet, 6 mázsa alatt veszteséges a termelés. Mekkora esélye van jövőre az Aranykalász kft.-nek arra, hogy veszteséges évet zár, s mekkora arra, hogy nyereséget hoz a földje?  
(Az átlag és a korrigált szórás kiszámítandó!)
- 2, Egy vállalat embereket keres felvételre. A jelentkezők átlagos életkora 38 év, a szórás 14 év. A vállalatvezetés burkolt preferenciája a 20–30 éves intervallumra korlátozódik. Mi az esély arra, hogy a jelöltek közül valaki e kategóriába tartozik (normál eloszlást feltételezve)?
- 3, Egy városban az egy főre eső jövedelem átlagosan 150 ezer forint, a szórás 30000 forint. Minek van nagyobb esélye (normáleloszlást feltételezve), hogy valaki 140–160 ezer forint között keres, vagy hogy 160 ezer forint felett keres? Mekkora esélye van 160 ezer forintnál kisebb fizetésre?  
(Figyelem! A 140–160 ezer forintos intervallumnál a görbe alatti területek különbsége lesz a megoldás, a 160 ezer feletti kategóriánál a táblázatból leolvasott eredményt ki kell vonni 1-ből/100%-ból!)
- 4, A Pétervásárai Homokkő Formációból álló lejtők átlagos lejtése 25%, a szórás 5%. Normáleloszlást feltételezve mekkora eséllyel találunk 33%-nál meredekebb lejtőt?

## **10. Matematikai statisztika – hipotézisvizsgálat – illeszkedésvizsgálat**

### **10. 1. Hipotézisvizsgálat**

A *hipotézis* (feltételezés) a mindennapi életünkben gyakran előfordul. Feltételezünk valamit és megvizsgáljuk, hogy a hipotézisünk helyes volt, vagy nem. A tudományos kutatásokban is alapvető szerepe van a hipotéziseknek. A hipotézist elméleti megfontolások, korábbi kutatási eredmények, vagy mért standard értékek alapján fogalmazzuk meg. Az empirikusan már vizsgálható állításokat kutatási hipotéziseknek nevezzük. Az olyan állítások, amelyek valamilyen

számított értékekkel, a változók minőségével kapcsolatosak, a statisztikai hipotézis témakörébe tartoznak. A hipotézisvizsgálat tehát egy statisztikai döntési feladat. El kell döntenünk, hogy az adott, a feladatra jellemző minta alapján mely állítást fogadjuk el helyesnek.

A kiinduló állításunkat mindig *nullhipotézis*nek ( $H_0$ ) nevezzük. Egy hétköznapi feltételezésnél tulajdonképpen mindegy, hogy mit választunk kiinduló helyzetnek (pl. valószínű hazudott, vagy nem valószínű, hogy hazudott), mert a hipotézisvizsgálattal kiderül az „igazság”. Abban az esetben viszont, ha tényleges számértékekkel dolgozunk, és döntenünk kell, akkor pontos valószínűségkalkulációt kell végezni. Valószínűségeket akkor tudunk pontosan számolni, ha az értékeinket csak a véletlen befolyásolja. Ezért tényleges számolások esetén nem mindegy, hogy mi a  $H_0$ . Ilyenkor a mérésünk eredménye csak a véletlentől (pl. műszer hibája) függhet. Alternatív hipotézisről akkor beszélünk, ha a nullhipotézis nem igazolódik be.

Először végezzünk egy hétköznapi hipotézisvizsgálatot. Tétélezzük fel, hogy január 15-én reggel van és a meteorológus jelentésében azt hallom, hogy Debrecenben a napi maximális hőmérséklet eléri a  $+12\text{ }^\circ\text{C}$  értéket. Mivel álmos voltam és sok egyéb tennivalóm is volt, ezért úgy gondoltam, hogy *valószínűleg* nem jól hallottam, ugyanis januárban nem szoktak ilyen magas hőmérsékleti értékek lenni. Ennek alapján az első feltételezésem az, hogy *nem valószínű*, hogy jól hallottam a  $+12\text{ }^\circ\text{C}$ -os értéket. Tehát gondolatban állást foglaltam.

Ez nem igényelt komolyabb mérlegelést, mert ebben a helyzetben el tudtam dönteni, hogy a  $+12\text{ }^\circ\text{C}$ -os értéket *valószínűnek*, vagy *nem valószínűnek* tekintem. Tehát tudom, hogy a *valószínű* és a *nem valószínű* között mi a határ.

Délben szépen sütött a Nap és az utcai digitális hőmérő  $12,5\text{ }^\circ\text{C}$ -ot mutatott. A hőmérséklet ismeretében a gondolatmenetem harmadik lépésében azt mérlegeltem, hogy ez mennyire tér el (van távol) a hipotézisemtől, amely szerint nem valószínű a  $+12\text{ }^\circ\text{C}$ -os napi maximum januárban. Állást kell foglalnunk, hogy *ekkora eltérést* kicsi vagy nagy valószínűségűnek tekintek („*valószínű*” vagy „*nem valószínű*”).

Megállapítom a tény, tehát „tévedtem”, tehát mégis jól hallottam a  $+12\text{ }^\circ\text{C}$ -os értéket. A negyedik lépésben tehát a valószínűség alapján döntöttem a hipotézisről, hiszen a tény ellentmond a hipotézisemnek.

A fenti gondolkodási lépések alkalmasak arra, hogy ismert tények alapján valamely ismeretlen helyzetről (hipotézisről) döntést hozzunk. Erről meggyőződhetünk, ha az előző feladatot a hipotézisünk ellentétével is megoldjuk. Az ellentétes hipotézisből kiindulva ugyanazon következtetésre jutunk. Tehát az eredmény nem függött attól, hogy mi volt az eredeti feltevésünk (hipotézisünk), a gondolatmenet jól működött, valóban kielégítette a tudományos követelményeket.

A hipotézis vizsgálatoknál első- és másodfajú hibákat követhetünk el.

		<b>H<sub>0</sub> hipotézist</b>	
		<b>elfogadjuk</b>	<b>elvetjük</b>
<b>H<sub>0</sub> hipotézis</b>	<b>Igaz</b>	Helyes döntés	<i>Elsőfajú hiba</i>
	<b>Hamis</b>	<i>Másodfajú hiba</i>	Helyes döntés

Leggyakrabban azt vizsgáljuk, hogy valamilyen statisztikai adatsorból kiválasztott rész-adatsor jellemzői (pl. várhatóérték, szórás, sűrűségfüggvény) eltérnek-e egymástól, vagy - valamilyen módon mért adatsor eltér-e egy ideális eloszlástól, illetve az adatsorok empirikus sűrűségfüggvényei, számtani közepei, szórásai eltérnek-e egymástól.

A nullhipotézisnél ( $H_0$ ) a különböző csoportba tartozó események, mérések matematikai jellemzői között nincs eltérés. Ha a  $H_0$  teljesedési valószínűsége nagyon kicsi, akkor a nullhipotézist elvetjük, és az eltérést szignifikánsnak nevezzük. Azt a valószínűséget, amelynél kisebb valószínűséget kapva a hipotézist elvetjük szignifikancia szintnek nevezzük. Értéke lehet tetszőleges, de általában a 0,05 és 0,01-et szokás választani.

A hipotézisek teljesülésének valószínűségét statisztikai próbák segítségével határozzuk meg. A tudományos hipotézisvizsgálat (statisztikai próba) valamilyen statisztikai függvényhez kapcsolódik, amelynek eloszlását a nullhipotézis ( $H_0$ ) határozza meg. Ha választunk egy „ $p$ ” valószínűségi szintet, akkor a statisztikai függvény eloszlásának ismeretében kiszámítható a függvénynek az a kritikus tartománya, ahova csak „ $p$ ” valószínűséggel esik a függvény értéke. A statisztikai próbát a következő lépésekben végezzük:

- Kiszámítjuk az adott minta alapján a választott statisztikai függvény értékét.
- Ha ez az érték a kritikus tartományba esik, akkor a hipotézisünk nem látszik igaznak, tehát a nullhipotézist elvetjük.
- Ha a statisztikai függvény értéke nem esik a kritikus tartományba, akkor a mintánk alapján  $1-p$  valószínűséggel hipotézisünk helyesnek látszik.

Természetesen előfordulhat, hogy az „ $p$ ”-t rosszul választottuk és így a döntésünk hibás lehet, vagy éppen azért döntöttünk hibásan, mert elfogadtuk a nullhipotézist. A hiba az „ $p$ ” változtatásával javítható. A döntés általában a kutatón múlik. Értelemszerűen, ami 80%-os valószínűséggel igaz ( $p=0,2$ ), az nem biztos, hogy 95%-os valószínűséggel is az lesz ( $p=0,05$ )



A hipotézisvizsgálat menete a következő:

- a hipotézis ( $H_0$ ) megfogalmazása,
- kiválasztjuk a megfelelő statisztikai próbát,
- megállapítjuk, hogy mi a határ a kis és a nagy valószínűség között (a döntésnél mit fogadunk el),
- megállapítjuk a szignifikancia szintet („ $p$ ”),
- végrehajtjuk a számolást,
- meghozzuk a döntést.

A különféle típusú hipotézisek vizsgálatára sokféle statisztikai próbát dolgoztak ki. A statisztikai próbák gondolatmenetében vannak közös vonások, ugyanis a statisztikai hipotézisvizsgálatok alapelve a nullhipotézisre épül, amelynek vizsgálatára általában az indirekt matematikai bizonyítás logikáját alkalmazzák. Ennek alapján pl.

- feltesszük, hogy nincs semmiféle kapcsolat az adatok között, akkor ez a nullhipotézis ( $H_0$ )
- elvégezzük a megfelelő számításokat és az eredményt összehasonlítjuk a nullhipotézis fennállása esetén érvényes eloszlással
- ha a számításaink eredménye nagyon valószínűtlen, akkor elvetjük a nullhipotézist és az ezzel ellentétes ún. alternatív hipotézist ( $H_1$ ) fogadjuk el.

A földrajzban, meteorológiában a nullhipotézis tehát leggyakrabban az, hogy két esemény bekövetkezése között nincsen összefüggés. Azt a  $p$  valószínűséget, melynek nem teljesülése/túllépése esetén a nullhipotézist elvetjük, szignifikanciaszintnek nevezzük. Tehát, ha a választott szignifikanciaszint  $p=0,05$ , akkor  $1-p = 0,95$ , 95%-os valószínűséggel állíthatom, hogy a két esemény között van összefüggés.

A létező sok statisztikai próbák közül a továbbiakban csak hármat tekintünk át.

## 10. 2. Az egymintás $t$ (Student) próba

Normális eloszlású változóknál használjuk és a szórást a mintából becsüljük. A várható értékre vonatkozó hipotézisünket akarjuk ellenőrizni.

A nullhipotézis  $H_0 = u$

Az egymintás  $t$ -próba próbastatisztikája:

$$t = \frac{\bar{x} - u}{s / \sqrt{n}}$$

ahol

- $\bar{x}$  a vizsgált valószínűségi változó átlaga a mintában,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

- $s$  a vizsgált valószínűségi változó becsült szórása,
- $u$  az előre adott érték, amihez az átlagot viszonyítjuk ( $H_0$ )
- $n$  a minta elemszáma.

A szórást itt többnyire az  $s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$  képlettel becsüljük, ahol a

minta az  $\{x_1, x_2, \dots, x_n\}$  értékekből áll.

Ha a minta elemszáma kisebb, mint 30 ( $n < 30$ ), akkor a szórás helyett a korrigált szórással számolunk, melyet  $s$  helyett  $s^*$ -gal jelölünk. Ennek képlete

$$s^* = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}, \text{ és } t = \frac{\bar{x} - u}{s^* / \sqrt{n}}$$

A feladatok megoldásának a főbb lépéseit egy gyakorlati példán mutatjuk be.

### **Példa:**

A kávé csomagolásakor az egyik gépen 250 g kávét kell tölteni a zacskókba. A gép mérőberendezésének pontatlansága miatt az egyik tömege kicsit több, a másik pedig kevesebb, mint 250 g. Arra vagyunk kíváncsiak, hogy az adagoló átlagos "teljesítménye" 250 g-nak mondható-e? A becsomagolt kávézacskókból választunk 12 db-ot, amelyeknek megmérjük a tömegét. A mérés eredményi: 247, 248, 252, 249, 247, 251, 249, 250, 248, 249, 250, 248.

A mérések átlaga  $\bar{x} = 249$ . A mérés után úgy gondoljuk, hogy nincs lényeges eltérés a tervezett 250 g-os értéktől.

Az adatsorból nehéz megállapítani, hogy lényeges eltérés van-e a 250 g-tól. Ennek az eldöntésére egymintás  $t$ -próbát alkalmazunk.

Feltesszük, hogy a kávé tömege, mint valószínűségi változó normális eloszlást követ. Ennek oka, hogy a mérési pontatlanságból eredő elérések általában a normáeloszlást követik. A tömeg g-ban való mérése arányskála, így az egymintás  $t$ -próba alkalmazásának feltételei teljesülnek. Mivel a minta elemszáma  $n = 12 < 30$ , így a szórás becsülésére az  $s^*$  képletet használjuk. Az  $u = 250$

- Az első lépés a  $t$  próbastatisztika értékének kiszámítása:

$$s^* = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{(247-249)^2 + (248-249)^2 + \dots + (248-249)^2}{12-1}} =$$

$$s^* = \sqrt{\frac{26}{11}} = 1.54 \quad t = \frac{\bar{x} - u}{\frac{s^*}{\sqrt{n}}} = \frac{249 - 250}{\frac{1.54}{\sqrt{12}}} = \frac{-1}{\frac{1.54}{3.46}} = \frac{-1}{0.45} \approx -2.22$$

- A második lépés a  $p$  szignifikancia szint megválasztása. Vegyük a szignifikancia szintet  $p = 0,05$ -nek, ami azt jelenti, hogy 5%-os kockázatot vállalunk arra, hogy esetleg úgy vetjük el a nullhipotézist, hogy az közben igaz.
- A  $p$  szignifikancia szinttől függő  $t$  érték kiválasztása a próbának megfelelő  $t$ -eloszlás táblázatból. A táblázat kétdimenziós, a  $p$  szignifikancia szint és az  $f$  szabadsági fok ismeretében azonnal megkapjuk a táblázatbeli  $t$  értéket. Az  $f$  szabadsági fokot az egymintás  $t$ -próba esetén az  $f = n - 1$  képlettel számítjuk. A szabadsági fok  $f = 12 - 1 = 11$ , így a  $p$  és az  $f$  ismeretében a  $t$ -eloszlás táblázatából\* könnyen kikereshetjük a megfelelő értéket, ami  $t_p = t_{0,05} = 2,201$
- Végül a nullhipotézisre ( $H_0$ ) vonatkozó döntés meghozása következik. Két lehetőség közül kell választanunk. Ha  $|t| \geq t$ , akkor a **nullhipotézist elvetjük**, viszont ha a  $|t| < t_p$ , akkor a **nullhipotézist megtartjuk**.

A jelenlegi feladatban  $|t| \approx 2,22$  miatt  $2,22 > 2,201 = t_{0,05}$  Ez azt jelenti, hogy a  $t$ -próba alapján a  $H_0$ -t elvetjük, mert a kávécsomagok súlya szignifikánsan eltér ( $p = 0,05$ -ös szignifikancia szint mellett) a 250 g-tól.

Ha a  $|t|$  értéke nem érte volna el a 2,201-es értéket, akkor az egymintás  $t$ -próba nem mutatott volna ki szignifikáns különbséget a vizsgált valószínűségi változó mintabeli átlaga és az adott  $u$  érték között (0,05 szignifikancia szint mellett).

Két összetartozó minta esetén a  $t$ -próba alkalmas a két minta összehasonlítására is akkor, ha vagy az  $x_i - y_i$  különbségek, vagy az  $x_i/y_i$  hányadosok normális eloszlásúak. Ha a feltételek teljesülnek, akkor a különbség esetén  $u = 0$ , hányados esetén pedig  $u = 1$  választással végezzük a próbát.

### 10. 3. A kétmintás $t$ próba

A kétmintás  $t$ -próba alkalmazhatóságának feltétele a szórások egyezése. Két adathalmaz (populáció) összehasonlítására szolgál átlagaik alapján. A két

\* a  $t$ -eloszlás táblázata a könyv végén a mellékletben található

adatsor adataiból meghatározott számtani közepek kisebb–nagyobb mértékben eltérnek egymástól. Az eltérés két okra vezethető vissza:

- Az adatok véletlenszerű változása eredményezi a különbséget.
- Az eltérés valóságos, két lényegesen különböző adatsorról van szó.

A kétmintás  $t$ -próba próbastatisztikája:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{(n-1)s_x^{*2} + (m-1)s_y^{*2}}} \cdot \sqrt{\frac{n \cdot m \cdot (n + m - 2)}{n + m}}$$

ahol

- $\bar{x}$  az egyik  $\bar{y}$  a másik valószínűségi változó átlaga a mintájában, tehát

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{és} \quad \bar{y} = \frac{1}{m} \sum_{j=1}^m y_j$$

- $s_x^*$  az egyik  $s_y^*$  a másik valószínűségi változó korrigált szórása,
- $n$  az egyik,  $m$  a másik minta elemszáma.

A feladatok megoldásának a főbb lépéseit egy gyakorlati példán mutatjuk be.

### **Példa:**

Egy meteorológiai állomáson két egymás utáni évben a táblázatban található havi csapadékadatokat mérték. Arra vagyunk kíváncsiak, hogy a két év csapadékadatai közötti különbség szignifikánsnak mondható-e, vagy nem. Feltételezzük, hogy az adatok között normális az eloszlás.

- Első lépésben ellenőrizzük, hogy a két adatsor szórása azonosnak tekinthető-e. Ezt külön statisztikai próba ( $F$ -próba) segítségével ellenőrizhetjük. Csak akkor alkalmazhatjuk a kétmintás  $t$ -próbát, ha az  $F$ -próba a szórások között szignifikáns különbséget nem mutat. Szignifikáns különbség esetén a kétmintás  $t$ -próbát nem lehet alkalmazni. Az  $F$ -próbához is a korrigált szórások négyzetét kell kiszámítani. (Nem szükséges feltétel a két populáció azonos mintaszáma,  $x$  és  $y$  különbözhet!)

	1. év		2. év	
	$x_i$	$x_i^2$	$y_i$	$y_i^2$
Január	24	576	36	1296
Február	23	529	5	25
Március	39	1521	54	2916
Április	35	1225	35	1225
Május	52	2704	66	4356
Június	112	12544	92	8464
Július	18	324	48	2304
Augusztus	89	7921	72	5184

Szeptember	48	2304	41	1681
Október	35	1225	30	900
November	56	3136	62	3844
December	37	1369	28	784
	$\sum x_i = 568$	$\sum x_i^2 = 35378$	$\sum y_i = 569$	$\sum y_i^2 = 32979$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{568}{12} = 47,3$$

$$\sum_{i=1}^n (x_i^2 - \bar{x}^2) = \sum_{i=1}^n x_i^2 - n \cdot (\bar{x})^2 \quad \text{alapján}$$

$$Q_x = \sum x_i^2 - n \cdot (\bar{x})^2 = 35378 - 12 \cdot (47,3)^2 = 8530,5$$

$$s_x^2 = \frac{Q_x}{n-1} = \frac{8530,5}{11} = 775,5$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{569}{12} = 47,4$$

$$Q_y = \sum y_i^2 - n \cdot (\bar{y})^2 = 32979 - 12 \cdot (47,4)^2 = 6018$$

$$s_y^2 = \frac{Q_y}{n-1} = \frac{6018}{11} = 547,1$$

$$F_{11,11} = \frac{s_x^2}{s_y^2} = \frac{775,5}{547,1} = 1,417.$$

A nagyobbik szórásnégyzetet osztottuk a kisebbik szórásnégyzettel, avagy a szórás a szórással. Az F indexében a számláló és nevező szabadsági fokát tüntettük fel. Az F-eloszlás kritikus értékei a  $p$  szignifikancia szinten  $p/2$  valószínűségeknél találhatók a táblázatban.

$$F_{(11,11)} = 3,48$$

$$F_{(11,11)} = 3,48 > F_{(11,11)} \text{ számított} = 1,417.$$

Tehát a két szórás közti különbség 95 % biztonsággal nem szignifikáns. Nincs eltérés. (Az F-próba Microsoft Excel program segítségével is megoldható!)

- A második lépés a  $t$  próbastatisztika értékének kiszámítása.

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{(n-1) \cdot s_x^{*2} + (m-1) \cdot s_y^{*2}}} \cdot \sqrt{\frac{n \cdot m \cdot (n+m-2)}{n+m}}$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{és} \quad \bar{y} = \frac{1}{m} \sum_{j=1}^m x \quad \text{képletek alapján} \quad \bar{x} = \frac{568}{12} = 47,3 \quad \bar{y} = \frac{569}{12} = 47,4$$

$$s_x^2 = 775,5, \quad s_y^2 = 547,1 \quad \text{és} \quad n=m=12$$

$$t = \frac{47,3 - 47,4}{\sqrt{11 \cdot 775,5 + 11 \cdot 547,1}} \cdot \sqrt{\frac{12 \cdot 12 \cdot (12 + 12 - 2)}{12 + 12}} = -\frac{0,1}{120,62} \cdot 11,49 = -0,01$$

- A szignifikancia szintet  $p = 0,05$ -nek választva és az  $f = n + m - 2 = 22$  szabadsági fok ismeretében a  $t$ -táblázatban a  $t_{0,05} = 2,074$  értéket találjuk, így A jelenlegi feladatban  $t \approx 0,01$  miatt  $0,01 < 2,074 = t_{0,05}$
- Mivel a  $|t|$  értéke nem érte el a  $2,074$ -es értéket, ezért a kétmintás  $t$ -próba ( $0,05$  szignifikancia szint mellett) nem mutatott ki szignifikáns különbséget a vizsgált két év adatai között.

#### 10. 4. A $\chi^2$ -próba

Több különböző típusú hipotézisvizsgálat tartozik ebbe a csoportba. Ezek a következők:

- Illeszkedésvizsgálat. Az illeszkedésvizsgálatban egy valószínűségi változó eloszlásának egy előre adott eloszlással való megegyezéséről döntünk, tehát a minta alapján tekinthető-e adott típusúnak a változó eloszlása, vagy nem. A nagy elemszám (legalább  $n > 10$ ) esetén általában a  $\chi^2$ -próbát alkalmazzák, mert ez alkalmas a folytonos és a diszkrét eloszlás vizsgálatára is.
- Függetlenségvizsgálat. A függetlenségvizsgálat során azt vizsgáljuk, hogy két mérési eljárásból származó adatsorok függetlennek tekinthetők-e vagy sem.
- Homogenitásvizsgálat. A homogenitásvizsgálat során azt vizsgáljuk, hogy a két mérési eljárásból származó adatsorból készített hisztogramok azonosnak tekinthetők vagy sem.

Az alábbiakban részletesen az illeszkedésvizsgálattal foglalkozunk.

#### Illeszkedésvizsgálat

Az illeszkedésvizsgálatokat szokták tiszta és becsléses típusba sorolni

- **tiszta illeszkedésvizsgálatról** akkor beszélünk, ha a feltevésünk egyetlen, teljesen meghatározott elméleti eloszlásfüggvényre vonatkozik, tehát a

gyakoriságokat ismert valószínűségekből kapott gyakoriságokkal hasonlítjuk össze.

- **becsléses illeszkedésvizsgálatot** végzünk, ha nem ismerjük annak az eloszlásnak a paramétereit, csak a típusát, amelyre a megfigyelt értékeket illeszteni szeretnénk. Az eloszlás típusa alapján a megfigyelt gyakoriságokból becsüljük az eloszlás paramétereit.  
A tiszta illeszkedésvizsgálat lehet egyenletes, vagy egyéb ismert paraméterű eloszlásra történő illeszkedésvizsgálat.
- **egyenletes eloszlásra történő illeszkedésvizsgálatról** akkor beszélünk, ha a feltételezett eloszlás egyenletes.  
A becsléses illeszkedésvizsgálat lehet normalitásvizsgálat, vagy egyéb becsült paraméteres illeszkedésvizsgálat.
- ha az eloszlás normális eloszlás, akkor **normalitásvizsgálatot** végzünk. Ilyenkor két paraméter becslésére van szükség.

A  $\chi^2$ -próba diszkrét és folytonos eloszlások vizsgálatára egyaránt alkalmas.

A próba menete:

A valószínűségi változó értelmezési tartományát osszuk fel  $m$  intervallumra (folytonos eloszlás esetén lehetőleg azonos hosszúságúra). Jelöljük az  $i$ -edik intervallumba eső mintaelemek számát  $f_i$ -vel.

$$\sum_{i=1}^m f_i = n$$

Az elméleti eloszlásfüggvény alapján határozzuk meg (becsléses esetben ennek paramétereit a mintából becsüljük) minden intervallumba esés  $p_i$  valószínűségét.

Az  $n \cdot p_i$  az  $i$ -edik intervallum elméleti eloszlása alapján várható gyakoriságát adja meg, ha a minta nagysága (az adatok száma)  $n$ . (Ettől a tapasztalati gyakoriság csak a mintavételi ingadozás hatására térhet el.)

A próbastatisztika számítását a  $\chi^2 = \sum_{i=1}^m \frac{(f_i - n \cdot p_i)^2}{n \cdot p_i}$  képlettel végezzük.

Ennek az értéke tiszta illeszkedésvizsgálat esetén  $(m - 1)$  szabadságfokú, becslés esetén  $(m - s - 1)$  szabadságfokú  $\chi^2$  eloszlást követ, ahol  $s$  a mintából becsült paraméterek száma.

Az eloszlásra vonatkozó hipotézisünket elvetjük, ha  $\chi^2 > \chi_\varepsilon^2$ , ahol  $\chi_\varepsilon^2$  az  $(1 - \varepsilon)$  valószínűségi szinthez tartozó táblázati érték.

**Példa:**

A fenti gondolatmenetet követve oldjuk meg a következő feladatot!

A földrajz szakos hallgatók félévi és évvégi zh eredményeit figyelembe véve határozzuk meg, hogy melyik félévben értek el jobb eredményt. Mindegyik alkalommal 90 fő irt dolgozatot. Egy-egy feladat helyes megoldása 5 pontot ért és az elérhető maximális pontszám 100 volt. Az elért eredményeket a következő táblázat tartalmazza:

elérhető pontszámok $x_i$	hallgatók száma félévkor $f_{i1}$	$x_i \cdot f_{i1}$	hallgatók száma év végén $f_{i2}$	$x_i \cdot f_{i2}$
45	2	90	1	45
50	4	200	3	150
55	7	385	6	330
60	8	480	8	480
65	8	520	11	715
70	10	700	13	910
75	12	900	10	750
80	16	1280	14	1120
85	15	1275	14	1190
90	5	450	6	540
95	2	190	2	190
100	1	100	2	200
$\sum_{i=1}^{12} =$	90	6570	90	6620

Feltételezzük, hogy a két félév eredménye között 0,05 szignifikancia szinten nincs jelentős eltérés ( $H_0$ ).

$$\sum_{i=1}^m f_{i1} = \sum_{i=1}^m f_{i2} = n = 90$$

$$\bar{x} = \frac{\sum f_{i1}}{m} = \frac{90}{12} = 7,5$$

$$Q_x = \sum x_i \cdot f_{i1} - m \cdot (\bar{x})^2 = 6570 - 12 \cdot (7,5)^2 = 5895$$

$$s_x^2 = \frac{Q_x}{12-1} = \frac{5895}{11} = 535,9$$

$$\bar{y} = \frac{\sum f_{i2}}{n} = \frac{90}{12} = 7,5$$

$$Q_y = \sum x_i \cdot f_{i2} - m \cdot (\bar{y})^2 = 6620 - 12 \cdot (7,5)^2 = 5945$$

$$s_y^2 = \frac{Q_y}{n-1} = \frac{5945}{11} = 540,5$$



$$F_{11,11} = \frac{s_x^2}{s_y^2} = \frac{535,9}{540,5} = 0,991$$

$$F_{(11,11)} = 3,48 > F_{(11,11)} \text{ számított} = 0,991$$

Tehát a két szórás közti eltérés 95 % biztonsággal nem szignifikáns, azok azonos alapsokaságból származóknak tekinthetők.

A feladatbeli **kapott** és a nullhipotézis ( $H_0$ ) igaz volta esetén **várt** gyakoriságokat hasonlítjuk össze és a köztük lévő különbségekből egy  $\chi^2$  próbastatisztikát számítunk.

$$\chi^2 = \sum_{i=1}^g \frac{(\text{kapott}_i - \text{várt}_i)^2}{\text{várt}_i}$$

szabadságfok:  $f = (\text{sorok} - 1) \cdot (\text{oszlopok} - 1) = (12-1) \cdot (2-1) = 11$

**Kontingencia táblázat elkészítése**

I.	II.	$\Sigma$
(2+1)·90/180=1.5	(1+2)·90/180=1.5	3.0
(4+3)·90/180=3.5	(3+4)·90/180=3.5	7.0
(7+6)·90/180=6.5	(6+7)·90/180=6.5	13.0
(8+8)·90/180=8.0	(8+8)·90/180=8.0	16.0
(8+11)·90/180=9.5	(11+8)·90/180=9.5	19.0
(10+13)·90/180=11.5	(13+10)·90/180=11.5	23.0
(12+10)·90/180=11.0	(10+12)·90/180=11.0	22.0
(16+14)·90/180=15.0	(14+16)·90/180=15.0	30.0
(15+14)·90/180=14.5	(14+15)·90/180=14.5	29.0
(5+6)·90/180=5.5	(6+5)·90/180=5.5	11.0
(2+2)·90/180=2.0	(2+2)·90/180=2.0	4.0
(1+2)·90/180=1.5	(2+1)·90/180=1.5	3.0
90	90	180

A próbamutató kis  $v_{ij} = \frac{\sum_{i=1}^k f_{ij} \cdot \sum_{j=1}^m f_{ij}}{N} = \frac{\text{sorösszeg} \cdot \text{oszlopösszeg}}{\text{teljes létszám}}$

$$\chi_{emp}^2 = \sum_{i=1}^k \sum_{j=1}^m \frac{(f_{ij} - v_{ij})^2}{v_{ij}}$$

A khi-négyzet értéke:

$$\chi^2 = \sum_{i=1}^m \frac{(f_i - np_i)^2}{np_i} = 0.998$$

$$\chi_{emp}^2 < 19.7 = \chi_{0.05}^2$$

A két félévben elért eredmények közel azonos értékűnek tekinthetők.

- 2, A földrajz BSc szakra felvételizők között a fiú-lány arány ugyanakkora-e? Tudjuk, hogy nem ugyanakkora, de feltételezzük, hogy az arány egyenlő, tehát  $H_0$ : Fiú:0.5 Lány:0.5

2007-ben 98-an nyertek felvételt, akik közül 39 volt a fiú, és 59 a lány. (ezek a kapott gyakoriságok:  $n_i$ )

Ha  $H_0$  igaz lenne, akkor 98-ból 49-49 fiúra és lányra számítanánk (várt/elméleti gyakoriságok:  $v_i$ )

Minél nagyobb az eltérés a kapott ( $n_i$ ) és a várt ( $v_i$ ) gyakoriságok között, annál valószínűbb, hogy a nullhipotézis ( $H_0$ ) nem igaz.

Az eltérés egy lehetséges mértéke:

$$\chi^2 = (n_1 - v_1)^2/v_1 + (n_2 - v_2)^2/v_2$$

Ha igaz a  $H_0$  hipotézis, akkor ez khi-négyzet eloszlású,  $f = 1$  szabadságfokkal.

Megoldás:

	fiú	lány	összesen
<b>Kapott gyakoriság (<math>n_i</math>)</b>	<b><math>n_1=39</math></b>	<b><math>n_2=59</math></b>	<b><math>N=98</math></b>
<b>Várt gyakoriság (<math>v_i=Np_i</math>)</b>	<b><math>v_1=49</math></b>	<b><math>v_2=49</math></b>	<b><math>N=98</math></b>

$$\chi^2 = (39 - 49)^2/49 + (59 - 49)^2/49$$

$$f = 2 - 1 = 1$$

$$\chi^2 = 2.041 + 2.041 = 4.082$$

$$\chi^2 = 4,082 < 6,635 = \chi^2_{001} (f=1)$$

Az eredmény miatt a  $H_0$  hipotézist elutasítjuk, és a fiúk aránya szignifikánsan kisebb a lányokénál.

- 3, A földrajz és földtudomány BSc szakra felvételizők között van-e különbség a fiú-lány arányban?

Feltételezzük, hogy az arány egyenlő, tehát a nullhipotézis ( $H_0$ ):

Fiú (földr.)= Fiú (földtud.) és Lány (földr.)= Lány (földtud.)

Szak	Fiú	Lány	Összesen
Földrajz	42	48	90
Földtudomány	32	28	60
Összesen	74	76	150

Ha a nullhipotézis ( $H_0$ ) igaz lenne, akkor a két szakon a közös fiú arány kb. 74/150 lenne, tehát a várt fiú-gyakoriság a földrajz és földtudomány szakon:

$$v_{11} = 90 \cdot 74 / 150 = 44,4 \text{ és } v_{21} = 60 \cdot 74 / 150 = 29,6$$

A közös lány-arány:

$$v_{12} = 90 \cdot 76 / 150 = 45,6 \text{ és } v_{22} = 60 \cdot 76 / 150 = 30,4$$

A nullhipotézis ( $H_0$ ) igaz volta esetén a

$$\chi^2 = \sum \frac{(n_{ij} - v_{ij})^2}{v_{ij}}$$

$f = 1$  szabadságfokú khi-négyzet-eloszlást követ, így  $\chi^2 < 3,841$  esetén ( $H_0$ ) -t megtartjuk,  $\chi^2 \geq 3,841$  esetén pedig  $H_0$ -t 5%-os szignifikancia szinten elutasítjuk (A táblázat szerint:  $\chi^2_{0,05} = 3,841$ ).

## ***Feladatok***

### ***Egy mintás t-próba***

- 1, Állítható-e 95%-os biztonsággal, hogy a pesti gyors menetideje 2:57 perc, ha hétfőtől szombatig a következő eredményeket mérjük: 2:50, 3:00, 3:15, 2:49, 3:10, 2:59?  
(Figyelem! Az adatok órában és percben vannak megadva – át kell számítani őket!)
- 2, EU-s pénzekből egy kistérség átlagosan 600 ezer euróval részesedik. Kiválasztva 10 cirill kistérséget, rendre 600, 500, 750, 600, 520, 530, 540, 520, 540, 610 ezer eurót kapunk. 5%-os tévedési lehetőséget feltételezve, az átlagtól való eltérés a véletlennek tudható be, vagy a cirill kormány lopja az EU pénzét?
- 3, A hivatalos adatok szerint a Kese-patak átlagos pH értéke 7,3 volt az elmúlt 20 évben. Terepi méréseink a következő eredményeket hozták a bőrcserző üzem betelepülése után: 6,7; 6,6; 5,9; 7,2; 6,4; 7,0; 5,7; 6,7; 6,7? 95%-os valószínűséggel állítható-e, hogy megváltozott a patak pH-ja?
- 4, Állítható-e 99%-os biztonsággal, hogy a szittyósberki kistérségben nem változott a munkanélküliség, hogyha tudjuk, hogy 10 évvel ezelőtt a 10 éves átlag 7,5% volt, s azóta a következő értékeket mértük: 10,2%, 7,8%, 6,7%, 6,9%, 7,8%, 8,9%, 8,8%, 11%, 7%, 7%?
- 5, Egy kormány azon kórházak bezárását tervezi, ahol az átlagos ágykihasználtság 75% alatti. Egy kórházban a következő értékeket mérték: 68%, 74%,

69%, 76%, 74%, 79%, 81%, 77%, 75%, 69%, 85%, 84%. 95%-os valószínűséggel állítható-e hogy nincs szignifikáns különbség az elvárt érték és a számított átlag között?

## **2 mintás t-próba**

- 1,** A külföldi befektetővel rendelkező települések esetében a foglalkoztatottság rendre 55%, 53%, 48%, 56%, 49%, 51%, 39%, míg a külföldi tőke nélküli települések esetében ez 48%, 44%, 46%, 52%, 38%, 38%, 35%, 46%, 35%. Van-e összefüggés a külföldi tőke jelenléte és a foglalkoztatottak aránya között  $p=0,05$  esetén ? És 99%-os valószínűséggel állíthatjuk-e ugyanezt?
- 2,** 95%-os biztonsággal állítható-e hogy a csatornázottság befolyásolja a talajvíz állását, ha a csatornázott kistaludi régió településein a következő felszíntől számított talajvíz értékeket mértük: 1.6, 1.7, 2.4, 2.6, 2.7, 2.9 m, míg a csatornázatlan településeken ez rendre: 1.3, 1.7, 1.5, 1.7, 1.9, 2.1, 1.3, 1.5, 1.8, 2.0, 1.4 m?
- 3,** Egy vállalatnál a férfiak fizetése rendre 59, 75, 48, 66, 74, 70, 59, 62 E forint, a nőké 56, 59, 60, 64, 71, 60, 61, 48, 49, 45, 44, 51, 52, 48, 41, 40, 52 E forint. 95%-os ( $p=0,05$ ) ill. 99%-os biztonsággal ( $p=0,01$ ) beszélhetünk-e a vállalatnál nemi diszkriminációról avagy nem érdemes a bírósághoz fordulnunk (amely elutasítja a nem megalapozott kereseteket)?
- 4,** Egy 30 % meredekségű lejtőn 2006-ban a következő eróziós értékeket mértük (mm-ben, a Gerlach-féle üledékcsapdában felhalmozódva): 45, 56, 74, 56, 67, 48, 66. Ugyanazon évben a völgy másik oldalán lévő eltérő kitettségű azonos alapközeten kialakult, de csupán 20%-os lejtőn ezek az értékek rendre 34, 47, 50, 66, 32, 54, 44, 60. A lejtés és a kitettség befolyásolják-e az erodált anyag mennyiségét ( $p=0,05$ )?

## 11. A becslés

A napi gyakorlatban (terepen és laboratóriumban) rendszeresen méréseket végzünk, vagy mások által mért eredményeket értékelünk, illetve ezek alapján becsléseket végzünk. A mérési eredményekből számított közelítést becslésnek, a közelítés számszerű értékét pedig becslött értékének nevezzük. Abban az esetben, ha a méréseket „pontosan” végeztük, akkor a mérési eredmény és a becslött érték közötti eltérését becslött hibának nevezzük.

Az adatok statisztikai feldolgozásoknál előfordul, hogy csak egy értékre (pl. relatív szintkülönbség kb. 16 m), máskor pedig olyan intervallumra (pl. napi középhőmérséklet 20–25 °C) adunk becslést, amelybe nagy valószínűséggel a várható érték beletartozik. Az elsőt a statisztikában pontbecslésnek, a másodikat intervallumbecslésnek nevezik. A becslés vonatkozhat átlagra, értékösszegre, arányra, szórásra.

### 11.1. A pontbecslés

A pontbecslés a statisztikai becslés olyan egyszerű esete, amikor az ismeretlent egy hozzá hasonlóval közelítjük (becsüljük). Először a mérések átlagát meghatározzuk, majd ezt úgy tekintjük, mint a várható érték egy becslését, illetve a mérési eredmények szórását úgy, mint a sokaság szórásának egy becslését.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

A pontbecslésnél végeredményként egy értéket adunk meg (pl. a hőmérséklet maximuma 24 °C közelében várható.)



### 11.2. Az intervallumbecslés

Az intervallumbecslésnél az átlag és a szórás alapján két számot (a, b intervallumot) úgy választunk, hogy az ismeretlen várható értéke bizonyos valószínűséggel az (a, b) intervallumban található. A becslés megbízhatóságát a valószínűség értéke mutatja. Ez természetesen megközelítheti a 100%-ot. Az intervallumbecslésnél az intervallum végpontjait és a valószínűség értékét adjuk meg (pl. a hőmérséklet napi maximuma 98%-os valószínűséggel 23 – 25 °C között várható).



A *matematikai statisztika* a paraméterek becslésekor használja a szabadságfok fogalmát. A szabadságfokok száma a mérési eredmények számával szoros összefüggésben van. Ha a mérési adatsor „ $m$ ” adatból áll, és csak ezeket az értékeket használjuk a becslés meghatározásához, akkor azt mondjuk, hogy a mérési sorozat (minta) szabadságfokainak száma „ $m$ ” (pl. egy mérési sorozat középértékét a számtani középpel becsljük). Előfordul olyan eset, amikor az adatsorból a becslt értéket úgy határozzuk meg, hogy ahhoz egy másik, már ugyanebből a mérési sorozatból becslt értéket (pl. számtani középértéket) használunk fel, akkor annyi szabadságfokot kell levonnunk az eredetileg „ $m$ ” szabadságfokból, ahány becslt értéket az újabb becslt érték meghatározásához felhasználtunk. Egyébként, ha a szabadságfokok számát nem csökkentjük, akkor a becslt értékünk pontosságát a már meghatározott becslt érték becslési hibája is terheli.

Legalább kettő mérést kell végeznünk, ha a mintából a középértéket, és a szórást is becslni akarjuk. Ha csak a mérendő mennyiségre vagyunk kíváncsiak, akkor egy mérés elegendő. Ebben az esetben azonban a szórás becslését nem tudjuk elvégezni.

### 11.3. A legkisebb négyzetek elve

A matematikai statisztikában az eloszlási paraméterek valódi értékének becslésére különböző módszereket dolgoztak ki. A mérési eredményeink statisztikai feldolgozásakor gyakran a legkisebb négyzetek elvén alapuló módszert alkalmazzuk. A mérések középértékének becslésénél azt vesszük figyelembe, hogy az eredményeknek a becslt értékektől való eltéréseinél, a becslt hiba négyzetösszege minimális:

$$F = \sum_i^n v_i^2 = \min. \quad \text{és} \quad v_i = u_i - \bar{u}_j$$

A  $v_i$  az  $i$ . mérési eredményre vonatkozó becslt hiba; az  $u_j$  a keresett  $j$ -ik mérendő mennyiség becslt értéke ( $j=1, 2, \dots, m$ ) és  $n$  a mérések száma. Az  $i$  értéke folyamatos és független a mérendő mennyiség sorszámától ( $j$ -től).

Ha a normális eloszlású mérési adatainknál a legkisebb négyzetek módszerét alkalmazzuk, akkor tulajdonképpen a legnagyobb valószínűség (maximum-likelihood) elvének megfelelő műveletet végzünk.

Az eredményeink feldolgozása során egy olyan  $v_i$  ( $i = 1, 2, \dots, n$ ) javítási értékrendszert határozzuk meg, amellyel a mérési eredményeinket úgy javítjuk, hogy azok között az ellentmondások megszűnnek.

#### 11.4. A középérték becslése

A középérték becslésénél gyakran a számtani középértéket és a mérési eredmények sorozatának középső értékét (a mediánt) határozzuk meg.

Egyetlen mennyiségre végzett  $n$  számú mérés esetén a képlet az

$$F = \sum_i^n v_i^2 = \sum_{i=1}^n (u_i - \bar{u})^2 = \min.$$

függvénnyel adható meg, ahol  $\bar{u}$  az ismeretlen a keresett mennyiség becslött értéke és  $v_i = u_i - \bar{u}$ . Az  $\bar{u}$  értéke az egyszerű számtani közép. A  $v_i = u_i - \bar{u}$  egyenlőség figyelembevételével

$$\sum v_i = \sum u_i - \bar{u} \cdot n \quad \sum v_i = 0$$

#### 11.5. A szórás becslése

A szórás becslését a mérési eredményeink pontosságának meghatározására használjuk. Ha a mérési eredményeink középértékeitől való eltéréseket hibáknak tekintjük, akkor a szórás egyfajta közepes hibaértéket (középhiba) jelent. Megkülönböztetünk előzetes és utólagos középhibákat. Abban az esetben, ha ismert az adatok középértéke, akkor előzetes középhibáról beszélünk. Az előzetes középhiba a következő képlettel számítható:

$$\mu = \pm \sqrt{\frac{\sum_{i=1}^n \Delta_i^2}{n}} \quad (\text{Gauss képlete})$$

A középhiba ( $\mu$ ) kettős előjelű lehet (pl. egy szimmetrikus tartomány esetén). Előzetes hibaértékeket szoktak megadni a műszerek leírásánál. Ezt az értéket az új műszer hitelesítési méréseinek adatai alapján határozzák meg.

Akkor, ha a középértéket az egyszerű számtani középpel határozzuk meg, akkor pedig utólagos középhibáról beszélünk.

A szórás (utólagos) becslése egyetlen középérték esetén:

$$\mu = \pm \sqrt{\frac{\sum_{i=1}^n v_i^2}{n}}$$

Az *utólagos középhiba* meghatározása segít a méréseink pontosságának megbecslésében. Az utólagos középhiba meghatározását mindig a felhasználó végzi. Tehát egy laboratóriumi, vagy terepi méréssorozat eredményeinek feldolgozásakor az előzetes és utólagos középhibák összehasonlításából ellenőrizhetjük a mérések korrektségét.

## 11.6. A konfidencia-intervallum fogalma

Általában célravezetőbb, ha a becsült számérték (pontosított becslés) helyett a becsült értékektől plusz, illetve mínusz irányban egy - egy határoló értéket adunk meg. Az így meghatározott intervallum bizonyos valószínűséggel a valódi értéket magába foglalja. A határoló értékeket konfidencia-határoknak, a konfidencia-határok által közrefogott tartományt konfidencia-intervallumnak (megbízhatósági intervallumnak) nevezzük.

Ez a konfidencia-intervallum olyan tartomány, amelyre az egységhez tetszőlegesen közel eső  $\beta$  valószínűséggel (konfidencia-szint) állítható úgy, hogy ez az intervallum lefedi a paraméter ismeretlen valódi értékét. A mérendő mennyiség valódi értékére felírható konfidencia-intervallum:

$$\bar{u} - t_{\beta} \cdot \frac{\mu}{\sqrt{n}} \leq U \leq \bar{u} + t_{\beta} \cdot \frac{\mu}{\sqrt{n}}$$

ahol a  $t_{\beta}$  - az ún. Student-eloszlás táblázatából (ld. melléklet) a  $\beta$  konfidencia-szint és az  $f = n - 1$  szabadságfok függvényében kiválasztható együttható akkor, ha a mérési eredmények eloszlása normális.

A konfidencia-intervallum terjedelme a választott *konfidencia-szinttől* függ. Minél nagyobb ez a terjedelem, annál nagyobb a valószínűsége annak, hogy az intervallum lefedi a valódi értéket. Nincs sok értelme annak, hogy az intervallumot túl nagyra válasszuk. Az erősen leszűkített intervallum pedig növeli annak a kockázatát, hogy már nem fed le a valódi értéket, így a következtetésünk biztonsága is csökken. Ezért az intervallum határainak a megadása nagy odafigyelést és szakmai felkészültséget igényel (pl. úrfelvételek interpretációjánál az elektromágneses hullámtartományok határainak kiválasztása).

## 11.7. Az adatok súlyozása

A mérési gyakorlatban előfordulhat, hogy az egyes mérési eredmények azonos középértékű, de különböző szórású alapsokaságból származnak. (pl. egy terepi felmérésnél ugyanazt a területet két alkalommal különböző pontosságú műszerrel mérjük.) A két különböző mérőeszközökkel kapott mérési eredmények alapján kívánjuk becsülni a terület valódi értékét és az egyes mérési eredmények szórását.

A rendelkezésünkre álló két adatsorból természetesen egyszerű számtani középértékkel nem becsülhetjük a terület valódi értékét. A feladat megoldásához egy új, a szórás függvényében megadott mérőszámra van szükség, amely mintegy összekötő szerepet tölt be az egyes, különböző szórású mérési eredmények között. A mérési eredményekkel csak akkor tudunk számolni, ha ugyanolyan szórású adatokból származtatjuk.



Végezzünk  $n$  számú mérést, amelynek értékei:  $u_1, u_2, u_3, \dots, u_n$ . A mérési eredmények szórása:  $\sigma_1^2, \sigma_2^2, \sigma_3^2, \dots, \sigma_n^2$ . Válasszunk olyan  $c = \sigma_0^2$  számértéket, amelyre  $c = p_1 \cdot \sigma_1^2 = p_2 \cdot \sigma_2^2 = p_3 \cdot \sigma_3^2 = \dots = p_n \cdot \sigma_n^2 = \sigma_0^2$  teljesül.

A továbbiakban ezt úgy tekintjük, mint a keresett azonos szórású alapsokaság szórásnégyzetét. Ez az összefüggés akkor áll fenn, ha a  $p_i$  ( $i = 1, 2, \dots, n$ ) szorzókat úgy választjuk meg, hogy azok

$$p_1 = \frac{\sigma_0^2}{\sigma_1^2}; p_2 = \frac{\sigma_0^2}{\sigma_2^2}; p_3 = \frac{\sigma_0^2}{\sigma_3^2}; \dots; p_n = \frac{\sigma_0^2}{\sigma_n^2};$$

legyenek. A  $p_i$  ( $i = 1, 2, \dots, n$ ) szorzót súlynak nevezzük. A súly a szórásnégyzettel fordítottan arányos mennyiség. Tehát a súly azt fejezi ki, hányszor megbízhatóbb, pontosabb az egyik mérési eredmény a másikkal.

A szórásnégyzet valódi értékét általában nem ismerjük, ezért a számítás kezdetekor a súlyokat a gyakorlatban az előzetes középhiba képletével számítjuk:

$$p_i = \frac{\mu_0^2}{\mu_i^2};$$

A  $\mu_0^2$  értéke lényegében tetszőlegesen választható, megválasztásában csak a súlyok matematikai kezelhetősége játszik szerepet.

A mérési eredményekhez különböző szórások és különböző súlyok tartoznak. Ezért a számtani közép, továbbá az ennek alapján becsült utólagos középhiba számításához az eredeti mérési eredményeinket úgy kell módosítanunk, hogy azok mindegyike éppen a  $\sigma_0$  ( $\mu_0^2$ ) szórású alapsokaságból származzon. Utóbbit úgy érjük el, ha mindegyik mérési eredményt osztjuk a hozzá tartozó szórással (előzetes középhibával) és szorozzuk  $\sigma_0$  ( $\mu_0$ ) - lal:

$$u'_1 = \frac{\mu_0}{\mu_1} \cdot u_1; u'_2 = \frac{\mu_0}{\mu_2} \cdot u_2; u'_3 = \frac{\mu_0}{\mu_3} \cdot u_3; \dots; u'_n = \frac{\mu_0}{\mu_n} \cdot u_n$$

Ez a transzformáció olyan mérési eredményekhez vezet, amelyeknek azonos középértéke és szórása van. Figyelembe véve a súlyra adott összefüggést, a fenti képletek az

$$u'_1 = u_1 \sqrt{p_1}; u'_2 = u_2 \sqrt{p_2}; u'_3 = u_3 \sqrt{p_3}; \dots; u'_n = u_n \sqrt{p_n}$$

alakban is felírhatók.

Ezzel az átalakítással a különböző szórású mérési eredményeket úgy kezelhetjük, mintha azok ugyanabból az alapsokaságból származnának.

A súlyozott számtani közép képlete:

$$\bar{u} = \frac{\sum_{i=1}^n p_i \cdot u_i}{\sum_{i=1}^n p_i}$$

$$\text{Figyelembe véve a } \sum_{i=1}^n p_i \cdot v_i = \sum_{i=1}^n p_i \cdot (u_i - \bar{u}) = \sum_{i=1}^n p_i \cdot u_i - \bar{u} \sum_{i=1}^n p_i$$

és a  $\sum_{i=1}^n p_i \cdot v_i = 0$  összefüggéseket a súlyozott számtani közép utólagos középhibája a következő képlettel felírható:

$$\mu_0 = \pm \sqrt{\frac{\sum_{i=1}^n v_i^2}{n-1}} = \pm \sqrt{\frac{\sum_{i=1}^n p_i v_i^2}{n-1}}$$

Mivel a mérési eredményeink szórása (középhibája) megegyezik, s a súly definíciója alapján minden  $u_i$  mérési eredmény súlya  $p_i = \frac{\mu_0^2}{\mu_i^2}$  azaz egységnyi, a

$\mu_0$  képlettel kifejezett értékét az egységsúlyúnak választott mérési eredmény utólagos középhibájának nevezzük (ez *elméletileg* megegyezik a tetszőlegesen választott előzetes középhibával). A súlyegység utólagos középhibájának és a súly definíciójának felhasználásával számíthatók az eredeti mérési eredmények utólagos középhibái.

$$\text{A } p_i = \frac{\mu_0^2}{\mu_i^2} \text{ képletből a } \mu_i = \frac{\mu_0}{\sqrt{p_i}}$$

ahol a  $\mu_0$  helyébe az utólagos értéket helyettesítjük. Azonos szórású mérési eredmények esetén természetesen minden súly egységnyi, ekkor nyilvánvalóan minden  $\mu_i$  - re igaz, hogy  $\mu_i = \mu_0$ .

### **Példák**

- 1, Egy mintaterület talajának pH értékét kell megbecsülnünk. Válasszunk ki a területről származó mintákból 10 db-ot. Ezek pH értékei: 7.52; 7.50; 7.56; 7.68; 7.52; 7.58; 7.48; 7.50; 7.69; 7.51  
A minták jellemzői:

$$\bar{x} = \frac{75,54}{10} = 7,554$$

$$s = \sqrt{\frac{\sum x_i^2 - n \cdot (\bar{x})^2}{n-1}} = \sqrt{\frac{341,75 - 10 \cdot 7,554^2}{10-1}} = \sqrt{\frac{570,68 - 570,6}{9}} = 0,0089$$

$$\bar{x} = \frac{1452}{10} = 145,2$$

$$s = \sqrt{\frac{\sum x_i^2 - n \cdot (\bar{x})^2}{n-1}} = \sqrt{\frac{401,6}{9}} = 6,68.$$

A  $t$  értéke 95%-os megbízhatósági szinten  $f = n - 1 = 9$  szabadsági foknál  $t_{95, \text{ táblázat}} = 2,262$ . A minták átlagos pH értéke  $|\mu|$  0,95 valószínűséggel

$$7,554 - 2,262 \cdot \frac{0,0089}{\sqrt{9}} \leq \mu \leq 7,554 + 2,262 \cdot \frac{0,0089}{\sqrt{9}}$$

$$7,55 \leq \mu \leq 7,56 \quad \text{intervallumba esik.}$$

A minták számának növelésével az intervallum szélessége csökken.

- 2, Térképészeti gyakorlaton a geográfus hallgatók 9 fős csoportjának minden tagja  $\pm 2$  cm-es szórású műszerrel méri a terepen kijelölt tereptárgy távolságát. Az a feladatunk, hogy a méterben megadott mérési eredmények felhasználásával becsüljük meg a távolságot (U).

A műszeres mérés eredményei: 148,36; 149,14; 148,52; 148,80; 149,04; 149,00; 148,60; 148,90; 148,92

A mérési eredményekből  $\bar{x} = \frac{\sum x_i}{n} = \frac{1339,28}{9} = 148,809$

Az objektum távolságának várható értéke 95%-os megbízhatósági szinten:

$$\bar{x} - t_\beta \cdot \frac{\mu}{\sqrt{n}} \leq U \leq \bar{x} + t_\beta \cdot \frac{\mu}{\sqrt{n}}$$

$$148,809 - 2,306 \cdot \frac{0,02}{\sqrt{9}} \leq U \leq 148,809 + 2,306 \cdot \frac{0,02}{\sqrt{9}}$$

$$148,79 \leq U \leq 148,82$$

A mérések számának növelésével a várható érték intervallumának szélessége tovább csökkenthető.



## 12. Dinamikus fizikai földrajz

A következőkben olyan fizikai földrajzi példákat mutatunk be, melyek a fluviális és általános geomorfológia, valamint a geológia alapkérdéseit helyezik jobb megvilágításba.<sup>1</sup>

### I. Szubvulkáni test kezdeti hűlési sebességét ( $v_T$ ) befolyásoló tényezők:

$$v_T = (\lambda \cdot T_0) / (c \cdot \rho \cdot H_L \cdot H_F)$$

ahol  $T_0$  a felszín el nem érő szubvulkáni test kiindulási hőmérséklete,  $c$  a szubvulkáni test fajhője,  $\lambda$  a fedőközet hővezetési együtthatója,  $\rho$  a fedőközet sűrűsége,  $H_F$  a fedőközet vastagsága,  $H_L$  a szubvulkáni test vastagsága.  $v_T$ -t °C/s-ban kapjuk, melyet érdemes °C/1000 évre átváltani a könnyebb kezelhetőség végett.

Tehát a lakkolit/sill hűlése egyenesen arányos a fedő hővezető képességével és a kiindulási  $T_0$  hőmérséklettel, míg minél vastagabb a szubvulkáni test és a fedőközet, annál lassabb a hűlés. A kapott érték kicsi, de még így is gyorsabb a hűlés, mint mélytörmzs, batolit esetén. Az utóbbi esetben a lassú hűlés (hiszen a fedőközet és maga a kőzettest is vastag, tehát mélyen van) lehetővé teszi nagy kristályok kialakulását, így holokristályos ekvigranuláris szerkezetű lesz a kőzet, míg egy szubvulkáni test esetében a gyorsabb hűlés miatt a kőzetüveg fog dominálni.

II. A szubvulkáni test hűlése exponenciális függvényt követ, aminek oka, hogy a mindenkor hűlési sebesség az aktuális hőmérséklettel egyenesen arányos, azaz a hűlés sebessége (az időegység alatt bekövetkező hőmérsékletcsökkenés) is csökken. A hűlés karakterisztikus ideje ( $\tau$ ) ennek lineáris közelítését jelenti, amikor a hűlés sebességét állandónak vesszük (a karakterisztikus idő tehát kisebb a ténylegesen mérhető hűlési időtől). Ebben az esetben a hűlés sebessége független a szubvulkáni test eredeti hőmérsékletétől. Képlete:

$$\tau = c \cdot \rho \cdot H_L \cdot H_F / \lambda$$

III. A gravitációs anomália (az elvárt gravitációs érték és a mért érték különbsége) és a domborzat ismeretében lehet meghatározni, hogy mekkora lesz a terület izosztatikus kiemelkedése vagy süllyedése. Az anomália oka, hogy a Föld

---

<sup>1</sup> A képletek részletes levezetése megtalálható Szunyogh G.: Dinamikus földrajz c. könyvében, valamint Szénás Gy.: Geofizikai teleptan c. munkájában.

alakja geoid, eltér a forgási ellipszoidtól, melyre a  $g$  meghatározható, nem homogén tömegeloszlású, és élénk domborzattal jellemezhető.

$$\Delta g = \gamma \cdot 2 \cdot \pi \cdot \rho_{si} \cdot \Delta H$$

Ahol  $\Delta g$  a gravitációs anomália (izosztatikus maradékanomália),  $\gamma = 6,67 \cdot 10^{-11}$  gravitációs állandó,  $\rho_{si}$  a sima (asztenoszféra, a kéregnél nagyobb sűrűségű plasztikus anyag amelyen a litoszféralemezek úsznak) sűrűsége,  $3300 \text{ kg/m}^3$ , a  $\Delta H$  pedig a várható kiemelkedés vagy süllyedés, melyet tehát csak az asztenoszféra sűrűsége és a gravitációs anomália befolyásol. Például orogenezisnél, hegységképződésnél  $1000 \text{ m}$  anyag felgyűrődése csak kisebb kiemelkedést jelent, mert közben a gyökérrégió belesüpped az asztenoszférába, egy hegység  $1000 \text{ m}$ -es magasságsökkenéséhez pedig a külső erőknek több mint  $1000 \text{ méter}$  anyagot kell elszállítani, hiszen az erózió beindultával a hegység, ahogy könnyebbedik, úgy folyamatosan ki is emelkedik. Eljegesedésnél, illetve a jégtakaró elolvadásakor hasonló esemény játszódik le: az Antarktisz kontinentális táblája pl. a jég súlya miatt a tengersizint alá süllyedt, míg a jégsapkától megszabadult Skandinávia az izosztázia miatt emelkedik (miközben pusztul is, hiszen a külső erők, mint pl. a víz munkavégző-képessége növekszik, ha nő a magasságkülönbség).

Ez átvezet minket 3 további példatípushoz: a hegységek maximális magasságának megállapításhoz, a víz munkavégző-képességének kiszámításához és a hegységek pusztulásához az idő függvényében.

**IV.** A hegység talpánál mérhető nyomás kiszámolható, ha a hegység súlyát ( $F_s$ ) elosztjuk az alapterülettel. Ha a hegység alapterülete  $A$ , magassága  $h$ , sűrűsége  $\rho$ , akkor a fektetett, háromszög alapú hasábként tekintett hegységünk súlya:

$$F_s = m \cdot g = V \rho \cdot g = A \cdot h/2 \cdot \rho \cdot g, \text{ a talpnál mérhető nyomás pedig } p = h/2 \cdot \rho \cdot g.$$

Ennek az értéknek nem szabad túllépnie a hegységet alkotó anyag törési szilárdságát ( $\sigma$ ). Azaz a hegység magasságának felső határa a

$$h/2 \cdot \rho \cdot g < \sigma \text{ képletből számítható.}$$

A gránit sűrűsége  $2,7 \text{ g/cm}^3$ , törési szilárdsága  $150\,000\,000 \text{ N/m}^2$ . Ebből következően a gránithegység maximális magassága  $11100 \text{ méter}$  körül lehet.

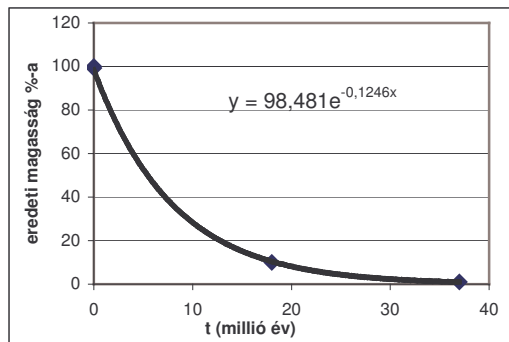
**V.** A denudáció (lepusztulás) számítására több modell létezik. Példánkban az egyik legegyszerűbből indulunk ki. Egy gránitos hegység egymillió év alatt

tömegének 10%-át veszíti el. Tételezzük fel, hogy a hegység téglalap alapú hasáb alakú, így a tömegvesztéssel azonos a magasság csökkenése is: gránit esetén 10%. Kiemelkedés nélküli modellt használunk, azaz 1 millió év múlva az eredeti magasság 9/10-e marad meg. Ha a denudációs rátát az egyszerűség kedvéért állandónak tekintjük, a következő évmillió során ismét ez játszódik le:  $0,9 \cdot 0,9=0,81$ . Tehát a képlet 10%-os veszteség esetén  $H_{akt}=0,9^t$  (ahol  $t$  millió évben van kifejezve). A hegység magassága hány millió év múlva csökken az eredeti egytizedére?

$$\begin{aligned} 0,1 &= 0,9^t \\ \lg 0,1 &= \lg 0,9^t \\ \lg 0,1 &= t \cdot \lg 0,9 \\ t &= \lg 0,1 / \lg 0,9 \\ t &= 22 \text{ (millió év)} \end{aligned}$$

Puhább kőzetben a veszteség nagyobb is lehet évmilliónként. Ha pl. egy mészkőhegységben a denudáció *emelkedés és bevágódás nélkül* millió évenként 20%-kal csökkenti a hegység eredeti reliefjét, úgy 11 millió év után még 10%-a, 22 millió év után még 1%-a marad meg.

Az itt kifejtetteket módosítja a tehercsökkenés miatti izosztatikus emelkedés, és ez esetben 10% 18 millió év



után, 1 % pedig 37 millió év után maradna (ábra). Ez egybevág azzal, amit jelenleg a hegységek létezéséről tudunk emelkedési főfázisuk lezárulta után. Az ábra a hegység pillanatnyi magasságát mutatja bármely  $t$  időpillanatban az eredetihez képest (empirikus formula a fenti példaadatok alapján).

Más denudációs egyenletek a relief (abból kiindulva, hogy a magasságkülönbség növeli a helyzeti energiát, és a végezhető munkát) vagy a lejtőszög alapján számítják a denudációt (Ahnert), negligálva a többi befolyásoló tényezőt, pl. kőzetkeménységet:

$$D=1,535 \cdot 10^{-4} \cdot h - 0,011 \text{ (m/1000 év)}$$

( $D$ : denudáció méterben,  $h$ : átlagrelief = a területegységen - 20x20 km<sup>2</sup> - lévő legalacsonyabb és legmagasabb pontok magasságkülönbségének középértéke)

$$D=0,967 \cdot \sin a - 0,007 \text{ (m/1000 év) (a: közepes lejtőszög)}$$

$$\text{ahol } \sin a = 0,005 + 0,00015 \cdot h$$

( $h$ : átlagrelief= a területegységen lévő legalacsonyabb és legmagasabb pont magasságkülönbségének középértéke)

Ezeknél a képleteknél a bemenő paraméterek felvett értékei is folyamatosan változnak, így bonyolultabb sorozatoknak tekinthetjük őket matematikai szempontból. Például a  $D = \sin a$  képlet esetében denudáció csökkenti a magasságot, így viszont a lejtőszög is megváltozik, a következő azonos művelet (**iteráció**) során tehát már a módosított lejtőszöget kell használni. Mivel a képletből következően a lejtőszög csökkenni fog hosszú távon, így a denudáció maga sem egyenletes. A denudáció előrejelzése (és egyéb, iterációt igénylő problémák megoldása) pl. az **Idrisi Macro Modeler**-ben lehetséges.

Egyes szerzők különbséget tesznek a felszín érettsége (idő) között:

$$D=4,19 \cdot 10^{-4} \cdot h-0,25 \quad (\text{Pinet-Souriou, fiatal orogének})$$

$$D=0,63 \cdot 10^{-4} \cdot h-0,02 \quad (\text{Pinet-Souriou, idős orogének})$$

Ahol  $h$  = magasság,  $D$ = denudáció (mm/év, v. m/1000 év)

Vannak kombinált denudációs egyenletek (pl. Scheidegger), melyek a kőzetkeménység és a lejtés függvényében számítják a felszínalacsonyodást (az izosztatikus kiemelkedést mellőzve):

$$D = -(1-c) \cdot \operatorname{tga} \cdot (1+\operatorname{tg}^2 a)^{1/2}$$

( $D$ : a felszín alacsonyodása egységnyi idő alatt,  $\operatorname{tg} a$ :  $dy/dx$  a két koordináta-érték,  $c$ : a kőzet keménysége,  $0$ : puha  $\leq c \leq 1$ : nagyon kemény)

**VI.** A folyóvíz munkájának energiamérlege a következőképpen írható fel: az összmunkavégzés (mely az energiaállapot megváltozását is jelzi, 0 súrlódást feltételezve) a hordalékszállítási és eróziós munka összegeként írható fel, ha a sebesség ( $v$ ) állandó.

$$W_{\text{össz}} = \Delta E_h = m \cdot g \cdot \Delta h = W_{\text{szállítási}} + W_{\text{eróziós}}$$

$m \cdot g \cdot h = \rho_{\text{víz}} \cdot V \cdot g \cdot h$ , ahol a  $V$  térfogat felírható  $L \cdot B \cdot H$  téglatestként ( $H$  a vízfolyás mélysége,  $B$  a vízfolyás szélessége,  $L$  a vizsgált szakasz hossza)

$\Delta h$  magasságváltozás lejtőn történik, a víz valódi úthossza ( $\Delta s$ ) tehát nagyobb, mint a függőleges  $\Delta h$ . Az út a lejtőszög sinusával ( $\sin a$ ), a sebességgel ( $v$ ) és az idővel ( $\Delta t$ ) arányos ( $h/\Delta s = \sin a$ ,  $\Delta s = v \cdot \Delta t$ ,  $h = v \cdot \Delta t \cdot \sin a$ ).

Az összmunka tehát:

$$W_{\text{össz}} = \Delta E_h = \rho_{\text{víz}} \cdot L \cdot B \cdot H \cdot g \cdot \sin a \cdot v \cdot \Delta t$$

A hordalékszállítási munka egyenesen arányos a fent említetttel azonos térfogatú vízben lévő hordalék súlyával ( $m_{\text{hord}}$ ) és a szállítási úttal ( $\Delta s$ ,  $\Delta s = v \cdot \Delta t$ ):



$$W_{szállítási} = \mu \cdot m_{hord} \cdot g \cdot \Delta s = \mu \cdot \rho_{hord} \cdot V \cdot g \cdot \Delta s = \Delta s \cdot \rho_{hord} \cdot L \cdot B \cdot H \cdot g \cdot \mu,$$

azaz

$$W_{szállítási} = \mu \cdot \rho_{hord} \cdot L \cdot B \cdot H \cdot g \cdot v \cdot \Delta t \quad (\mu = 0,2 \text{ J/Nm})$$

Az erózió  $B$  szélességben,  $L$  hosszúságban játszódik le a mederfenéken bizonyos vastagságban ( $\Delta h_{erodált}$ ). Az eróziós munka arányos a térfogattal:

$$V_{erodált} = L \cdot B \cdot \Delta h_{erodált}$$

$$W_{eróziós} = \sigma \cdot V_{erodált} = \sigma \cdot L \cdot B \cdot \Delta h_{erodált}$$

A  $\sigma$  a fajlagos eróziós munka, mely Magyarországon  $10^{10}$ - $10^{12}$  J/m<sup>3</sup> a mederanyag keménységétől függően.

Mivel  $W_{összes} = W_{eróziós} + W_{szállítási}$ , így

$$\rho_{vz} \cdot L \cdot B \cdot H \cdot g \cdot v \cdot \Delta t \cdot \sin \alpha = \sigma \cdot L \cdot B \cdot h_{erodált} + \mu \cdot \rho_{hord} \cdot L \cdot B \cdot H \cdot g \cdot v \cdot \Delta t$$

Időegység alatt pedig (osztva  $\Delta t$ -vel):

$$\rho_{vz} \cdot L \cdot B \cdot H \cdot g \cdot v \cdot \sin \alpha = \sigma \cdot L \cdot B \cdot h_{erodált} / \Delta t + v \cdot \rho_{hord} \cdot L \cdot B \cdot H \cdot g \cdot \mu$$

Az erózió mértéke az időegység alatt lejátszódó medermélyülés  $v$  szélesedés, amely a fenti képlet átrendezésével:

$$v_{erózió} = h_{erodált} / \Delta t = (1 / \sigma) \cdot g \cdot H \cdot v \cdot (\rho_{vz} \cdot \sin \alpha - \rho_{hord} \cdot \mu)$$

A vízfolyás időegység alatt végzett összmunkája (a felszabadult helyzeti energia) egyenesen arányos az áramlási sebességgel, a meder lejtésével, a meder szélességével és mélységével.

A modell szerint a *hordalékszállítási munka* tehát egyenesen arányos a víz sebességével, a hordaléksűrűséggel, az egységnyi súlyú hordalék szállításához szükséges munkával ( $\mu$ ), a folyó szélességével és mélységével.

Az *eróziós munka* egyenesen arányos a kőzet keménységével ( $\sigma$ ), az erodált térfogattal (mélyülés és vízfolyásszélesség).

Az *erózió sebessége* egyenesen arányos a vízmélységgel, a sebességgel, fordítottan arányos a fajlagos eróziós munkával, tehát keményebb kőzeten lassabb az erózió. A hordaléksűrűség nagysága szintén csökkenti az erózió sebességét.

Ha a felszabaduló helyzeti energia nagyobb, mint a szállítási munka, akkor a különbség medermélyítésre (erózió) fordítódik. Egy Duna jellegű vízfolyásnál 1 km mederhossz esetén másodpercenként 7,5 millió J energia szabadul fel, ebből a hordalékszállítás 1,5 millió J, tehát medermélyítésre marad 6 millió J, ami 0,6 mm/év medermélyülés/mederszélesedés. Tehát ez km-enként 7,5 MW felszabadulását jelentené optimális esetben. Ha a felszabaduló helyzeti energia kisebb, mint a szállítási munka, akkor negatív erózió, azaz akkumuláció játszódik le, a folyó képtelen magával szállítani a hordalékokat.

**VII.** Az árhullám haladási sebessége kiszámítható a következő képlet alapján (a modell akkor érvényes, ha az árvíz nem hagyja el a medret, az ártérre nem lép ki, gátak között marad) :

$$c = v_0 + (g \cdot (H_m/H_0) \cdot (H_m/2 + H_0/2))^{1/2}$$

ahol  $c$  az árhullám terjedési sebessége lefelé a folyón,  $v_0$  a folyó eredeti sebessége,  $H_m$  az áradó folyó vízoszlopának magassága (vízállása),  $H_0$  pedig az árvíz előtti vízállás. Az árhullám terjedési sebessége tehát közvetlenül nem függ az árvízben áramló részecskék sebességétől, viszont az utóbbi kiszámításához ugyanazon paraméterekre van szükség. Az árhullám terjedési sebessége meghaladja az árvízben áramló részecskék sebességét, így az árhullám előre halad, gyorsabban, mint a víz mozgásából ez következne, veszélyességét éppen ez adja.

**VIII.** Az árvízben áramló víz sebessége kiszámolható:

$$v_a = v_0 + ((H_m - H_0)/H_m) \cdot (g \cdot (H_m/H_0) \cdot (H_m/2 + H_0/2))^{1/2}$$

Ne felejtjük el, hogy a helyzeti energia mozgási energiává alakul át ( $m \cdot g \cdot h = m \cdot v^2/2$ ), a víz pusztító ereje tehát sebességének négyzetével arányos, így a kisvízi munkát többszörösen felülmúlja. Egy átlagos  $1 \text{ m/s}$ -mal mozgó vízfolyás munkavégző-képessége tehát kilencszerese lesz, ha sebessége  $2 \text{ m/s}$ -mal nő ( $3 \text{ m/s}$  lesz), és ezt a munkát a szabályozások után az eredeti ártérnél jóval kisebb területen végzi, így a gátakra nehezedő nyomás is sokkal nagyobb (a kisebb terület és magasabb vízoszlop miatt), mint korábban a nagyobb területű ártéren. Ugyanakkor a kanyarulatok levágásával meggyorsult az árvíz lefolyása, így a torkolati szakaszokon kisebb valószínűséggel fordul elő, hogy két árhullám (pl. Körös-Tisza, Maros-Tisza) utóléri egymást és szuperponálódik hatásuk.

**IX.** Ha a vízszint-emelkedés elenyésző az eredeti vízszinthez képest (pl. több km mély tengeren), cunami (tengerrengések) esetén, amikor a nyílt tengeren a hullámmagasság alig  $1\text{-}2 \text{ m}$ , így nem különböztethető meg a szél okozta hullámzástól, vagy pl. tölcsértorkolatban felnyomuló dagályhullámkor (pl. pororoca), a képlet a következőképpen egyszerűsíthető:

$$c = v_0 + (g \cdot H_0)^{1/2} \text{ Ha a sebesség ismert, a mozgási energia számítható:}$$

$$E_m = m \cdot c^2/2, \quad \text{ahol } m = \rho_{\text{víz}} \cdot V, \quad V = L \cdot B \cdot H$$

## Feladatok

- 1, A gravitációs anomália (izosztatikus maradékanomália) egy területen  $3 \cdot 10^{-4} \text{ m/s}^2$ . Mekkora a várható izosztatikus elmozdulás, ha a SiMa sűrűsége  $3300 \text{ kg/m}^3$ ?
- 2, Mennyi egy 500 fokos lakkolit lehűlésének karakterisztikus ideje, ha a sűrűség  $2600 \text{ kg/m}^3$ , a lakkolit fajhője  $890 \text{ J/kg} \cdot ^\circ\text{C}$ , a fedőközet hővezetési együtthatója  $2 \text{ W/m}^\circ\text{C}$ , a lakkolit vastagsága  $300 \text{ m}$ , a fedőközeté  $200 \text{ m}$ ?
- 3, Mekkora a folyó eróziós és összes munkája továbbá a bevágás sebessége, ha a víz sűrűsége  $1000 \text{ kg/m}^3$ ,  $g=10 \text{ m/s}^2$ , a hordaléksűrűség  $0,7 \text{ kg/m}^3$ , a  $\mu = 0,2 \text{ J/Nm}$ , a  $\sigma = 10^{12} \text{ J/m}^3$ , a folyó mélysége  $6 \text{ m}$ , szélessége  $400 \text{ m}$ , a lejtés  $0,05^\circ$ , a sebesség  $= 1,2 \text{ m/s}$ , a vizsgálatot  $1000 \text{ m}$ -es folyóhosszon és  $1 \text{ s}$ -ra végezzük el?
- 4, Ha a folyó hordaléksűrűsége  $1,4 \text{ kg/m}^3$ , a lejtés  $0,005^\circ$ , a sebesség  $1 \text{ m/s}$ , mederkeresztmetszet  $100 \text{ m} \cdot 5 \text{ m}$ ,  $1000 \text{ m}$  hosszú vizsgálunk, ( $\sigma = 10^{12} \text{ J/m}^3$ ,  $\mu = 0,2 \text{ J/Nm}$ ,  $g = 10 \text{ m/s}^2$ , víz sűrűsége  $1000 \text{ kg/m}^3$ ) akkor  $t = 1 \text{ s}$  alatt mennyi a végzett munka (felszabadult  $E$ )? Mennyi fordítódik hordalékszállításra és mennyi bevágásra? Mekkora a bevágás mértéke egy év alatt?
- 5, Hányszorosára nő az árhullám terjedési sebessége és munkavégző-képessége (energiája), valamint a turbulens áramlással mozgó víz sebessége, ha  $1 \text{ m}$ -rel emelkedik a vízszint? (Eredeti vízszint  $5 \text{ m}$ , eredeti sebesség  $= 1 \text{ m/s}$ )  
(Figyelem! Az energia a sebesség négyzetével arányos, tehát pl. egy 5-szörös sebességnövekedés 25-szörös munkavégző-képességet eredményez!)
- 6, Számítsuk ki, hogy egy mészkőhegység maximális magassága mekkora lehet (sűrűsége  $2,3 \text{ g/cm}^3$ , törési szilárdsága  $100\,000\,000 \text{ N/m}^2$ ).
- 7, Kiemelkedés nélküli modellel számítva, ha egy mészkőhegység egymillió év alatt magasságának  $20\%$ -át veszíti el.  $5$  millió év múlva milyen magas lesz a hegység? Mikor éri el a hegység az eredeti magasság  $10\%$ -át?
- 8, Tengerrengés után a hullámok haladási sebessége  $3 \text{ m/s}$ , a vízoszlop magassága  $2000 \text{ m}$ . Számítsuk ki a cunami haladási sebességét a nyílt vízen és mozgási energiáját ( $1 \text{ m}$  széles fronton,  $1 \text{ m}$  hosszon)!
- 9, A  $d = \sin a$  ( $d$ : a denudáció mértéke méterben,  $1000$  év alatt,  $a$  lejtőszög) képlet alapján határozza meg az eredetileg  $300$  méteres relatív relieffel rendelkező  $25$  fokos lejtő magasságát és lejtőszögét és a lepusztult anyag arányát az eredetihez képest:  $1000$  év,  $100$  ezer év és egy millió év múlva (Idrisi Macro Modeller használata ajánlott)!



### 13. Korreláció- és regressziószámítás

A korreláció- és regressziószámítást a geográfiában gyakran használjuk két vagy több változó közötti kapcsolat irányának, szorosságának és jellegének a meghatározására. Ha csak arra vagyunk kíváncsiak, hogy van-e egyáltalán kapcsolat a változók között, akkor korrelációt számítunk. A kimutatható kapcsolat még nem jelent ok-okozati viszonyt!

Ha a nincs kapcsolat két  $(x, y)$  változó között, tehát ha  $x$  nem befolyásolja  $y$  értékét ( $y$  értéke független  $x$  értékétől), akkor korrelálatlan a kapcsolat. Sztochasztikus kapcsolatról akkor beszélünk, ha  $x$  és  $y$  értékek között egyértelmű, függvényyszerű kapcsolatot nem lehet megállapítani, de az  $(x, y)$  adatpárok halmaza egy adott típusú – lineáris, vagy valamilyen görbe vonalú – függvénnyel közelíthető. Függvényyszerű a kapcsolat, ha minden egyes  $x$  értékhez egyértelműen egy-egy  $y$  értéket rendelünk hozzá.

A két változó között lehet laza, vagy szoros a függvénykapcsolat. Ha létezik kapcsolat, akkor regressziószámítással határozhatjuk meg, hogy az egyik változó értékeiből hogyan lehet előre jelezni a másik változó értékeit. A kapcsolat iránya lehet pozitív, vagy negatív. A kapcsolat jellege pedig lehet lineáris, vagy exponenciális. A korreláció és a regresszió között sok a hasonlóság, ha a korreláció mérőszáma (a korrelációs koefficiens) szignifikáns, akkor mindig szignifikáns lesz a lineáris regresszió is. A két változó összefüggésének vizsgálata esetén meghatározzuk a regressziós függvény paramétereit, a korrelációs együtthatót és az összefüggésre vonatkozó szignifikanciaszintet.

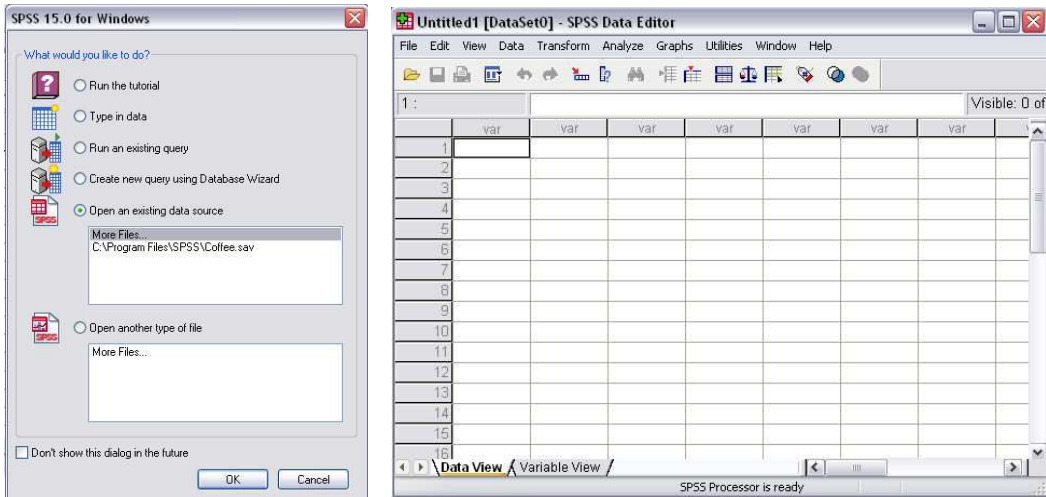
A statisztikai kézikönyvek általában határozott különbséget tesznek a korreláció- és regressziószámítás között. A feladatok megoldásához különféle statisztikai szoftvereket használnak. Mi az SPSS szoftvert választottuk, ezért szükségesnek tartjuk a szoftver rövid bemutatását. Elsősorban azokat a menüpontokat ismertetjük, amelyek a mintafeladatok megoldásához szükségesek

#### 13. 1. Az SPSS szoftver főbb jellemzői.

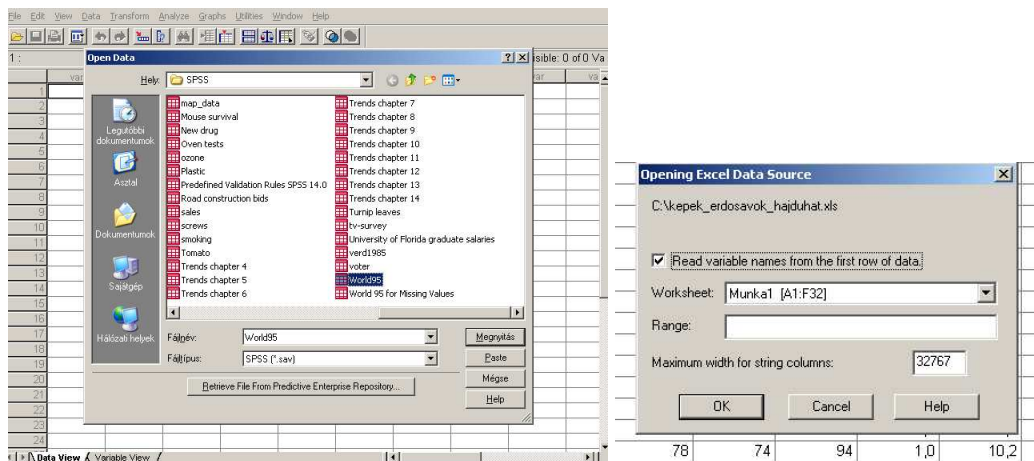
Az SPSS program indításakor két párbeszédablak jelenik meg egymáson. A felső ablakban tanulmányozhatjuk a program leírását, adatokat vihetünk be, illetve a korábban elmentett adatbázisokból válogathatunk. Természetesen ezt az ablakot be is zárhatjuk, mert ezeket a műveleteket a második ablakban a program menüpontjaiból is elérhetjük.

Az SPSS olvassa a saját fájlformátumait (.sav és .spo – az utóbbi az üzeneteket, műveleteket és eredményeket tartalmazza SQL-nyelvben), továbbá az Excel (xls) és DBASE (dbf) fájlokat is. Az Access (mdb) fájlok tábláit használat előtt xls-be kell exportálni. A megnyitás a **File-Open-Data** menün keresztül lehetséges. Figyeljünk arra, hogy az Excel file esetén a használni kívánt munkala-

pot állítsuk be a worksheet legördülő menüjében, illetve ha van fejlécezés az oszlopainknak, változónevekként menthetjük őket. Az adatbeolvasáson túl adatbevitelre is lehetőségünk van, ekkor vegyük figyelembe, hogy az SPSS inkább az Access adatbáziskezelőre hasonlít, mint az Excelre, így adatok kijelölésekor, másolásakor érhetnek meglepetések.



Az SPSS indításakor megjelenő párbeszédablakok



A sav és xls kiterjesztésű file-ok megnyitása File/Open menüponttal

Az SPSS-ben két nézet van, a **Data View** és a **Variable View**. Az előbbi az adatbázist tartalmazza, az utóbbi az adatbázis szerkezetére vonatkozó informá-

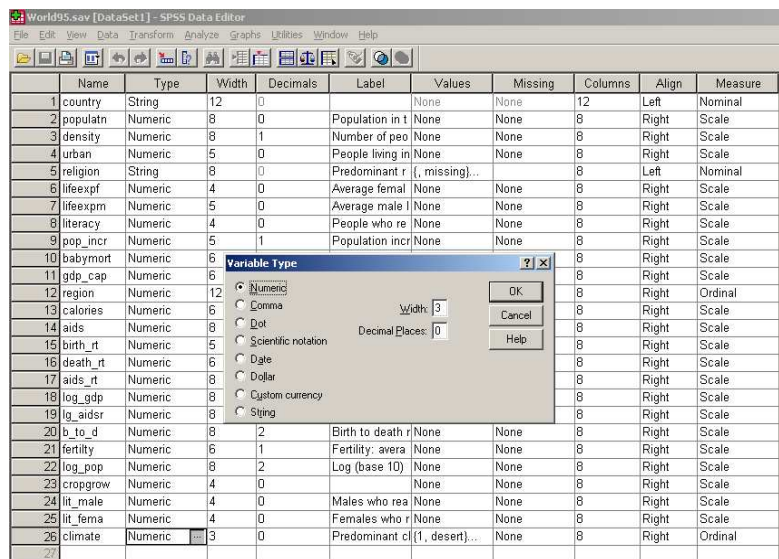
ciókat. Az előbbi tehát oszlopokból (változók) és sorokból (esetek) áll, az utóbbi metaadatokat tartalmaz.

A **Variable View**-ban van lehetőségünk beállítani a változó elnevezését, a változó típusát (leggyakrabban string vagy numeric), az oszlop karakterszélességét, a decimálisok (tizedeshelyek) számát. Excelből történő beolvasás esetén előfordulhat, hogy tizedestörteink egész számmá alakulnak, ekkor nem történik adatvesztés, és az eredeti számértékek helyreállíthatók a változók típusának, vagy a decimálisok számának átalakításával. Lehetséges a rövid változónevekhez, a jobb azonosítás érdekében megjegyzést fűzni, és itt találhatóak a változók beállítási: a skála, ordinális és nominális adatok, melyek szintén átállíthatók. (A rossz beállítás hibaforrás lehet).

Lehetőségünk van új változót is definiálni (értelemszerűen adatokkal nekünk kell feltölteni a **Data View**-ban). Ha bármelyik műveletünk során hibát jelez a program, vagy az utasítást végre sem hajtja, érdemes először a **Variable View**-ban lévő beállításokban keresni a hiba okát.

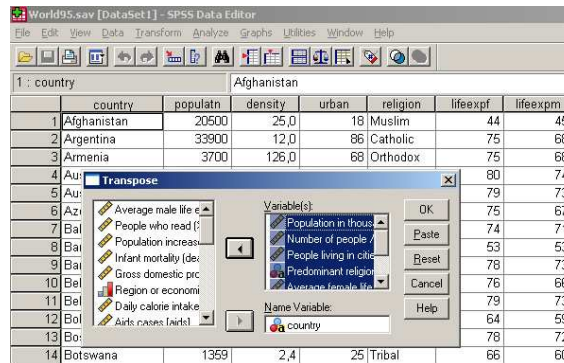
A **Data View Edit** menüpontja segítségével lehetőség van sorok és oszlopok beszúrására, törlésére, sorba rendezésre, de ezek a fejlécen, illetve az eseteket tartalmazó első szürke oszlopon történő jobb kattintásra is előjönnek (**clear, insert variable, insert case, sort ascending, sort descending**).

**Find** és **Goto case** paranccsal kereshetjük meg a kérdéses kifejezéseket (abban az oszlopban keres a program, ahol a kurzor áll) rekordokat. Fontos felhívni a figyelmet arra, hogy a keresés csak egész kifejezésekre vonatkozik, rekordon belüli karakterekre nem.



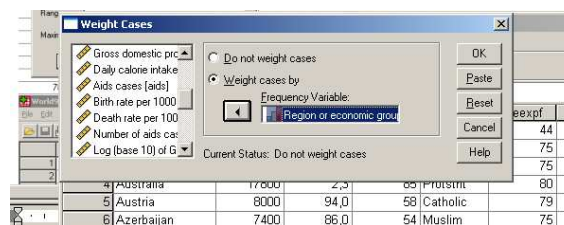
Adatbeállítás SPSS-ben

Előfordulhat, hogy az adatszerkezet megkívánja, hogy a sorokból oszlopokat, az oszlopokból sorokat csináljunk, ebben az esetben a **Data - Transpose** menüt használhatjuk. Itt ki kell választani azt a változót, amiből oszlopféjléceket akarunk csinálni (**Name variable**).



CASE_LBL	populatr	Alghanistan	Argentina	Armenia	Australia	Austria	Azerbaijan	Bahrain	Bangladesh	Barbados	Belarus	Belgium	Bolivia
1 populatr	20500,00	33900,00	3700,00	17800,00	8000,00	7400,00	600,00	125000,00	256,00	10300,00	10100,00	?	
2 density	25,00	12,00	126,00	2,30	94,00	86,00	628,00	600,00	605,00	50,00	329,00		
3 urban	18,00	86,00	68,00	85,00	58,00	54,00	83,00	16,00	45,00	65,00	96,00		
4 religion													
5 lifeexpf	44,00	75,00	75,00	80,00	79,00	75,00	74,00	53,00	78,00	76,00	79,00		

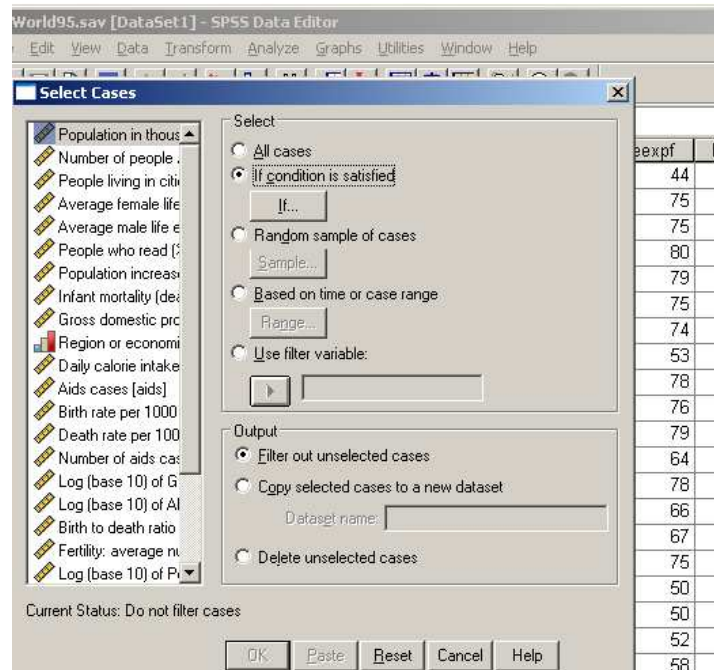
A **Data - Weight cases** menü súlyozott számításokat tesz lehetővé, kiválaszthatjuk azt a változót, mellyel a többit súlyozni akarjuk. Ilyen lehet például egy közterület térfogata (mint változó) és sűrűsége (mint súlyfaktor) - ebben az esetben a réteg (eset) tömegét tudjuk kiszámolni, de ilyen lehet a völgyek irányultsága (változó) és hossza (súlyozó tényező). Ne felejtsük el visszavonni a súlyozást, ha már nincs rá szükség.



A **Data - Select Cases** sokoldalú válogatást, szűrést tesz lehetővé az adathalmazban. egyik lehetőség az **if** (ha) szerkezet használata, azaz feltétel megadása, ebben az esetben a kifejezőszerkeztő segítségével egy sor egyszerű és bonyolultabb logikai és halmazműveletet végezhetünk el az oszlopokkal (pl. na-



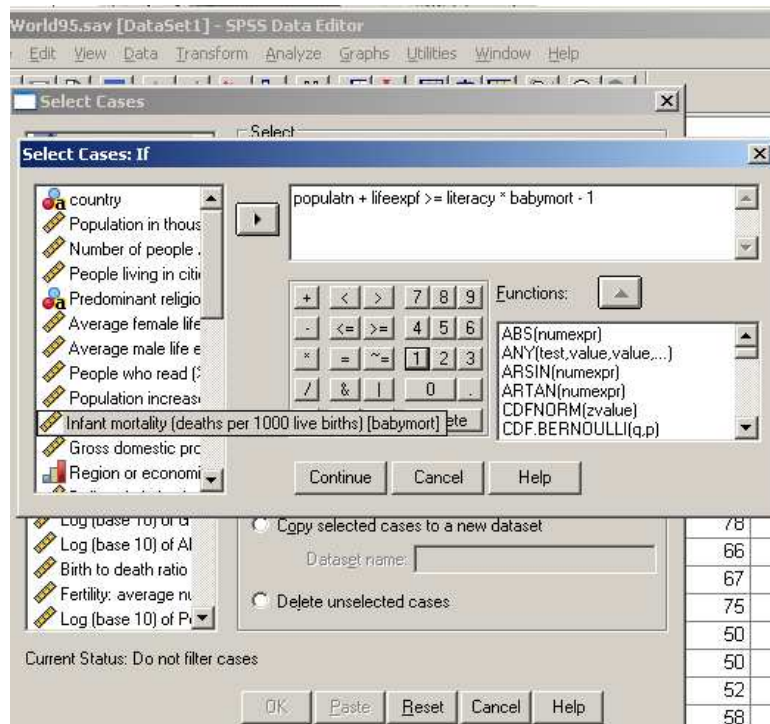
gyobb mint, kisebb mint, nem egyenlő, „a változó” + „b változó” > „c változó”–5 stb.). A válogatás (mintavételezés) szempontjait nekünk kell megtervezni, hogy logikailag helytálló legyen – azokat az adatokat válogassuk ki, amit valóban akarunk. A halmazból adatok törölhetők (**Delete unselected cases**), vagy kiszűrhetők (**Filter**). Az utóbbi esetben az így kiszűrt esetek nem törölődnek, de nem is vesznek részt a további műveletekben, ezt az első oszlopban egy átlós áthúzás jelzi. Az előző esetben a sorok, rekordok törölődnek (de ha nem mentjük el a file-t az eredeti nevén, akkor a program bezárásakor a módosítások nem mentődnek el). Az is lehetséges, hogy a kiválogatott adatokat egy új file-ba mentjük, ebben az esetben az eredeti adathalmaz integritása szintén nem sérül.



Az adatok leválogatásának másik módja a véletlenszerű kiválasztás (**Random**). Ebben az esetben megadhatunk %-os értéket, vagy darabszámot. A harmadik módszer az intervallum megadása (**Time or case range**). Itt csak egy intervallum megadása lehetséges. Természetesen a három módszer kombinálható, azonban egyszerre csak egyet hajt végre a program, ezért mindig megfontolandó, hogy a sorrend felcserélése ugyanazt a végeredményt adja-e.

Egyes parancsok nem vonhatók vissza (az így fellépő problémák ellen csak úgy védekezhetünk, hogy kilépéskor nem mentjük a változtatásokat, vagy mindig az eredeti adatbázis másolatával dolgozunk), mások igen, így például a

sorok és esetek törlése. A visszavonás a legtöbbször csak a legutolsó műveletre terjed ki!

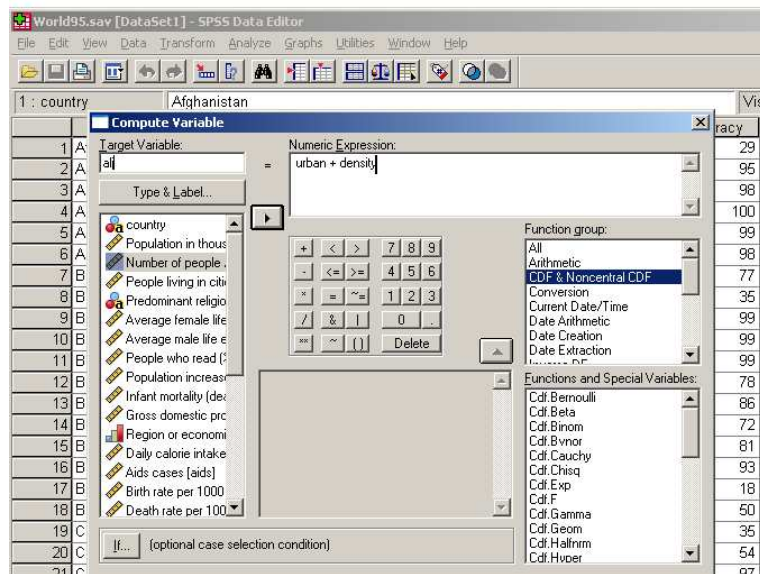


A **Transform** menüből a **Compute variable** az egyik legfontosabb parancs. Itt lehetséges az oszlopokkal (vagy a megszürt oszlopokkal) műveleteket végezni. A sorokkal matematikai művelet nem végezhető, pl. nem adható össze az értékük. A kifejezésszerkesztő segítségével vagy SQL nyelvben adhatók meg a végzendő műveletek. Ekkor meg kell adnunk a célváltozó nevét, amelybe az új értékek kerülnek. Célszerű új változót megadni, különben a program a már létezőt írja felül.

A műveletek közül a függvények, bonyolultabb kifejezések a jobb oldalon találhatóak. Szögfüggvényeknél fontos, hogy az adatokat radiánban kéri, tehát ha nem abban vannak megadva adattáblánkban, előtte át kell az értékeket számolni. A képletben ott, ahol egy változóra van szükség, egy „?” jelenik meg. A változókat a bal oldalról lehet összeválogatni. Alul látható a szűrésre lehetőséget adó **if** szerkezet, melynek működését már ismertettük. Ebben az esetben a végzett művelet nem kerül alkalmazásra az összes soron/eseten.

A másik fontos parancs a **Recode**. Érdekes új változóban kódolunk (**Recode into new variable**). Itt van lehetőség például áttérni szám típusú változóra szöveges változóról (vagy fordítva), vagy diszkrét értéket intervallumokba sorolni (csoportképzés). Ekkor meg kell adnunk a régi értékeket tartalmazó és az

új változó nevét egyaránt, majd a **Change** paranccsal aktiválni. Ezt követően választhatunk, hogy az egész adatbázison, vagy annak egy leválogatott részén végezzük el az átalakítást. Az **if** gomb itt is alul található. Majd meg kell határoznunk, hogy az egyes régi értékek milyen új értékeket kapnak.



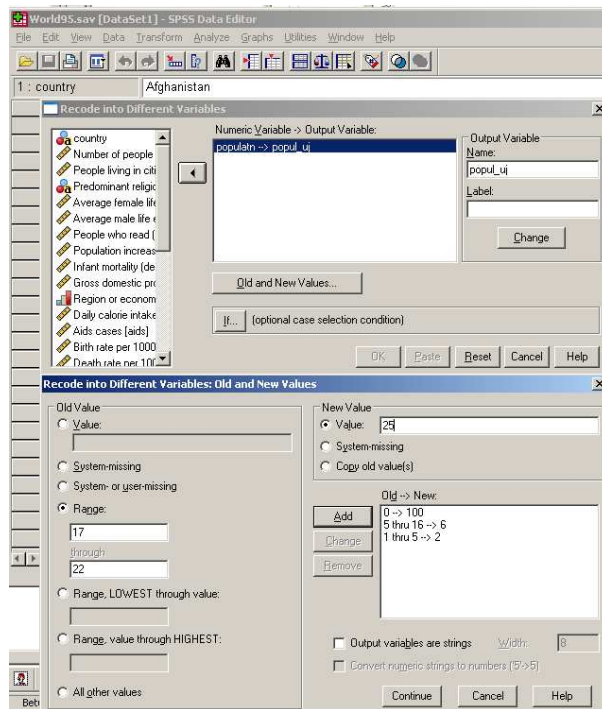
A régi értékek a bal, az új értékek a jobb oldalon vannak. Ha sok értéket kell újradefiniálnunk, akkor előtte célszerű róluk egy listát készíteni, hacsak nem akarunk folyton kilépni az alkalmazásból.

A régi érték megadható konkrét értéként (**Value**) és megadható intervallumként (**Range**), az új értékek között azonban intervallumot már nem adhatunk meg. Az újradefiniálást egyenként végezzük el.

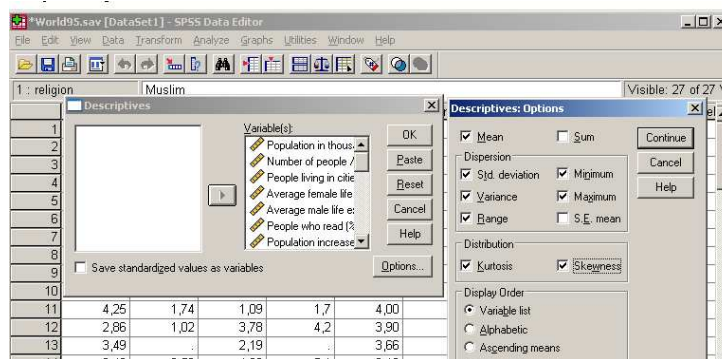
Az **Add** gomb megnyomásával kerül be az új érték a rendszerbe, ekkor ráterhetünk a következő átírt szándékozott értékre vagy intervallumra. Ha mégis módosítani kívánnánk az átkódolás szabályain, azt a **Change** ill. a **Remove** paranccsal tehetjük meg (előtte jelöljük ki a módosítandó értéket). Végül lehetőségünk van szöveges formátumként megadni az adatot (tehát az eddig számmal jelzett csoportot névvel ruházhatjuk fel – **Output variables are strings**), és a numerikus stringeket is valódi számmá lehet konvertálni (**Convert numeric strings to numbers**). Döntéseinket jól gondoljuk át mielőtt a **Continue** gombra, majd az **OK**-ra kattintunk.

Előfordulhat, hogy adathiány miatt nem teljes az adatsor, s bizonyos műveleteket nem végez el a program (pl. a klaszteranalízis), ha egy esetnél ismeretlen egy változó értéke. A hiányzó értékek helyett érdemes a csoportátlaggal számolni, bár ez adathamisításnak tűnhet, statisztikailag mégis ekkor a legkisebb a

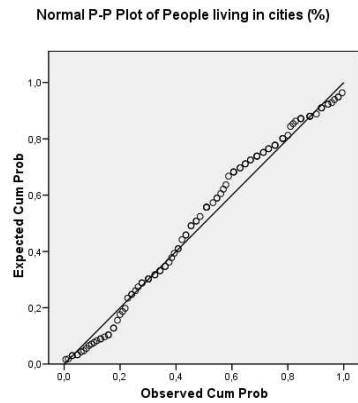
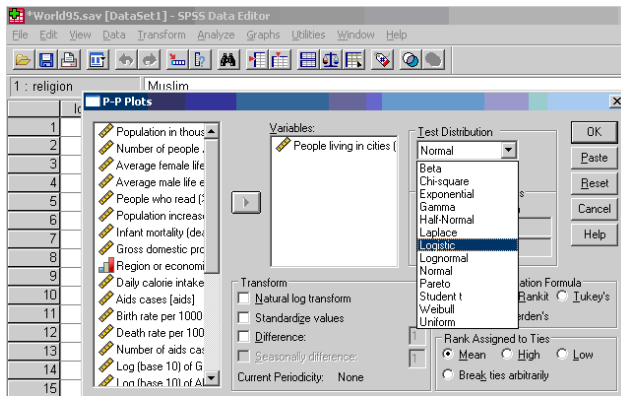
tévedés valószínűsége, ezt a program gyakran automatikusan felajánlja sok műveletnél.



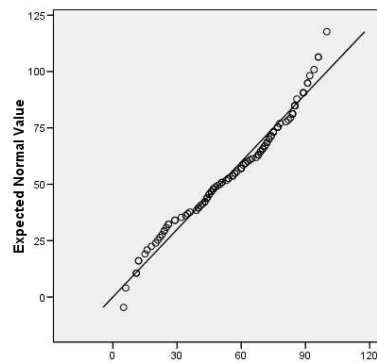
Az **Analyze** menüpontban található **Descriptive Statistics - Descriptives** menüt, mely az adathalmazok összehasonlítását segítő statisztikai paramétereket állítja elő. Az **Options** gomb megnyomásával a ferdeség, csúcsosság, terjedelem, szórás, átlag, összeg, variancia között választhatunk. Szintén elmenthetők változóként adathalmazunk standardizált értékei (erre pl. a klaszteranalízisnél lehet szükség). Ha több változót tesztelünk, akkor az adatok egy mátrixban jelennek meg. Az spo felületről az adatok kijelöléssel és **ctrl c** – **ctrl v** kombinációval wordbe és excelbe is átvihetők



Ugyanitt a **P-P plot** segítségével tesztelhetjük adatsorunk eloszlását. Kiválaszthatunk egy vagy több változót, beállíthatjuk a tesztelni kívánt eloszlást (az ábrán normáeloszlást választottunk, s láthatjuk, hogy az adatsor közelít a normáeloszláshoz, az „átló” közelében maradnak a pontok). Természetes alapú logaritmussal transzformálhatjuk az adatsor eloszlását, standardizálhatjuk a változót még a tesztelés előtt, ha nem bízunk az eredeti eloszlásban. Hasonló lehetőséget biztosít a **Q-Q plot** is.

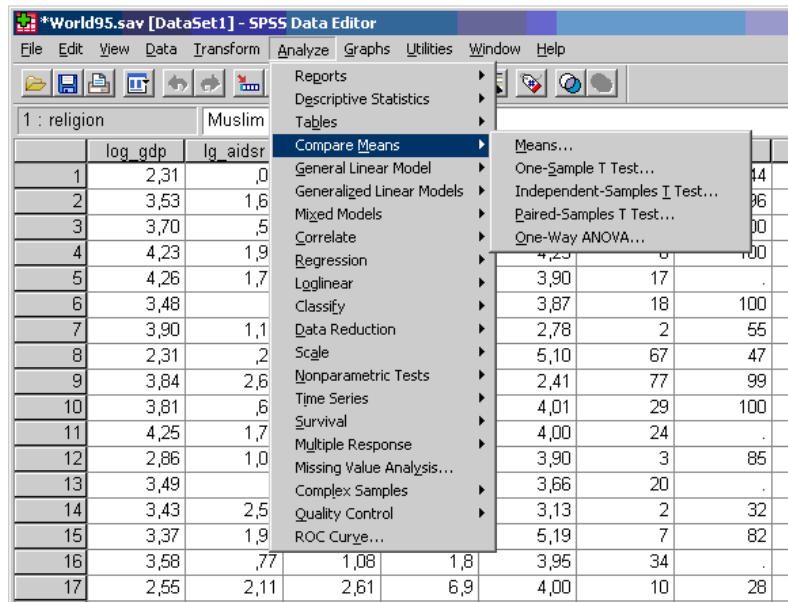


Normal Q-Q Plot of People living in cities (%)



A **Graph- Legacy Dialogs - Histogram** menüvel szintén vizsgálható egy adatsor eloszlása, amennyiben szükséges, az intervallumok kialakításában is segíthet a hisztogram. Az egymintás **Kolmogorov-Szmirnov-próba (Analyze - Nonparametric Tests - One Sample K-S)** szintén alkalmas eloszlások tesztelésére, részletes leírását lásd a korrelációs vizsgálatoknál.

Lehetőségünk van keresztábrák készítésére (**Crosstabs**). A **Compare Means** menüben belül végezhetünk egy mintás és 2 mintás t-próbát.



### *Disztancia–vizsgálatok (többdimenziós skálázás)*

Alkalmas valós és fiktív terek létrehozására a mintavételi pontok közötti távolság alapján.

Valós teret hozunk létre, ha pl. városok közötti közúti, vagy vasúti távolságokkal rendelkezünk, s ez alapján kell nagy mennyiségű adat térbeli elhelyezkedéséről képet alkotnunk.

Fiktív teret hozunk létre, ha például személyek, vagy politikai csoportosulások közötti viszonyt minősítünk pl. egy 1-től 10-ig terjedő skálán (a minősítő kritériumok megtervezése külön odafigyelést igényel, és saját módszertana van), majd megkíséreljük felrajzolni a társadalom mintázatát, belső csoportjait. De becült távolságok estében alkalmas ún. mentális térkép elkészítésére is, sőt az egyes „térképezők” tudatában élő torzulások, különbségek mérésére is.

Legegyszerűbben az SPSS segítségével, az **Analyse** menüpont **Multidimensional Scaling** almenüjének **Alscal** parancsával alkothatunk ilyen térképet. Fontos, hogy az adatokat már távolságként vigyük be az SPSS adatkezelőjébe (**Data are distances**). Az adatbevitelnél egy félmátrixot kell készíteni: a sorok és oszlopok fejlécében ugyanannak a kifejezésnek kell szerepelnie (pl. városok, személyek), s csak az átló alatti részeket kell kitölteni. Értelemszerűen az átlóban végig 1-esek (vagy nullák) szerepelnek, hiszen ua. kifejezések egymástól való távolsága 1.

## Két és háromváltozós pontdiagramok, mint koordinátarendszerek

Ebben az esetben nem a földrajzi tér dimenzióit ábrázoljuk, hanem matematikai/fiktív térben dolgozunk. 2, illetve 3 változót ábrázolhatunk, attól függően, hogy sík vagy térkoordináta-rendszerrel van szó. Több változó esetén – ha nem akarjuk kétváltozós diagramjaink számát túlzott mértékben megnövelni (10 változó esetében a kombinációk száma  $10!/2$ ) választhatjuk az adatredukciót:

- faktoranalízissel összevonjuk változóinkat,
- főkomponens-analízist végzünk és csak a fő komponenseket (mint a faktorokkal legnagyobb korrelációt mutató eredeti változókat) választjuk ki ábrázolásra
- összevonjuk a változókat (a+b és c+d vagy a/b és c/d lesz a két tengely).

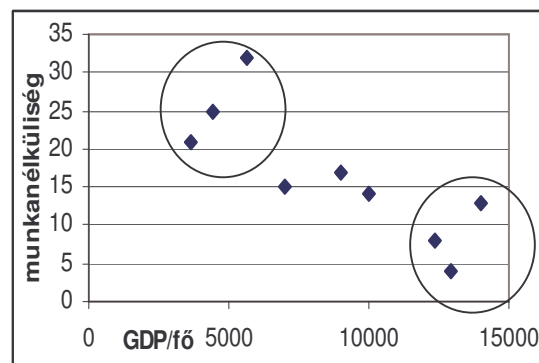
A kétváltozós pontdiagramot két alapvető célra használhatjuk:

- a**, csoportképzés esetén (minél jobban szétdobjuk az eltérő adatokat és minél közelebb kerüljenek egymáshoz az összetartozónak vélt adatok). A kétváltozós pontdiagram felfogható a klaszterképzés (lásd klaszteranalízis) legegyszerűbb változataként, ahol a dimenziók száma nem  $n$ , hanem csak 2.

A példaként bemutatott ábrán megfigyelhető, hogy a GDP/fő növekedésével csökken a munkanélküliség. Ez egyrészt nem általános érvényű törvényszerűség, másrészt az ok–okozati viszonyok megállapítására a módszer még a törvényszerűség megléte esetén sem alkalmas!

település	GDP/fő (USD)	munkanélküliség (%)
Pusztafalu	5620	32
Avasmérgecs	3650	21
Kishuta	4410	25
Aranyosmeggyes	12365	8
Dióskút	6984	15
Berecke	9985	14
Sártoronya	14002	13
Vályogvásári	8990	17
Erdőkertes	12898	4

Az adatok fiktívek

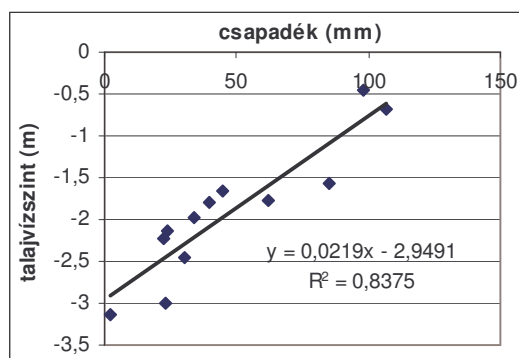


- b**, trendjelleg-vizsgálat esetén – ekkor a vizsgálati cél az, hogy az adatok elhelyezkedése a diagramon leírható-e egy függvénnyel (és ha igen, milyen pontosan). Minél jobban illeszkednek az adatok egy reg-

ressziós függvényre, annál erősebb lesz a kapcsolat a vizsgált változók között (mutatói:  $r$  korrelációs koefficiens és  $R^2$ , lásd a korrelációs és regressziós vizsgálatoknál részletesen)

A következő ábrán – mely felfogható egy olyan Descartes-féle koordináta-rendszernek, mely nem területi kiterjedést ábrázol – azt láthatjuk, hogy a havi csapadékmennyiség és a talajvízszint egy vonal köré rendeződik, azaz van függvénnyel leírható szabályszerűség közöttük, és a pontok illeszkedése a függvényre szintén vizsgálható.

hónap	vízszint (m)	csapadék (mm)	hőmérséklet
1	-2,14	24	-1
2	-2,22	22	0
3	-0,45	98	2,3
4	-0,69	107	5,6
5	-1,79	40	9,6
6	-1,66	45	16,7
7	-1,98	34	18,4
8	-3,14	2	27,8
9	-2,99	23	24,5
10	-1,56	85	15,2
11	-1,78	62	5,4
12	-2,45	30	1,6

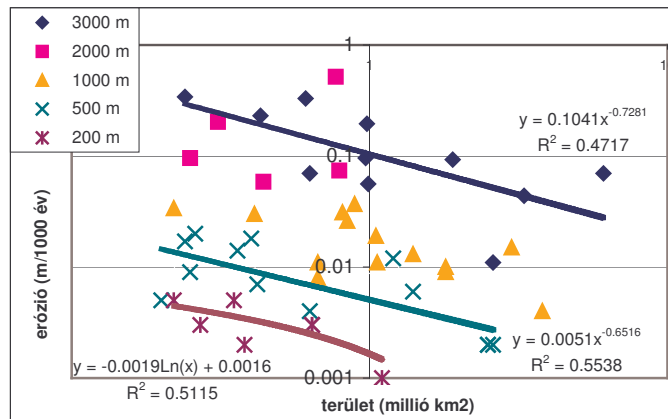


Az adatok fiktívek

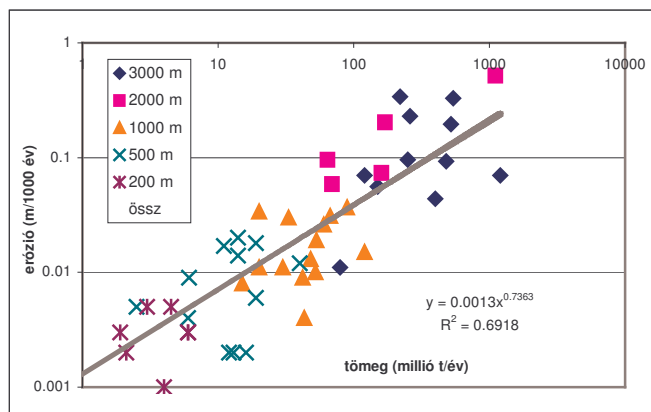
Az ábrázolás szempontjából az egyik változó, a hónapok sorrendje irrelevánssá válik a talajvíz és csapadékértékek ábrázolása esetén. Ugyanakkor szerepe nem elhanyagolható, hiszen az egyik változó értékeinek változása „késést” szenvedhet – itt értelemszerűen a beszivárgás ideje lesz a módosító tényező – mely befolyásolja a kapcsolat erősségét. Ekkor gyakran alkalmazott fogás, hogy az egyik változót „eltoljuk” a késleltetés értékével, tehát módosítjuk az értékpárokat.

Az is előfordulhat, hogy nincs látható trendjelleg az adatsor egészének ábrázolása esetén, de ha részekre bonjuk az adatsort, akkor mégis lehetséges a csoportosítás úgy, hogy trendjelleg rajzolódjon ki. Ehhez nagy tapasztalat és jó kérdésfelvetés kell:





Összefüggés a denudáció és a vízgyűjtő terület nagysága között 50, különböző magasságú régióban



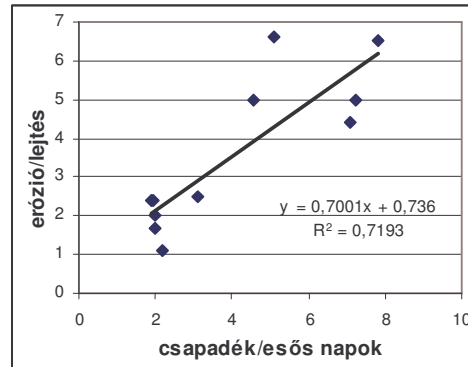
Összefüggés a szállított anyag össztömege és a denudáció között 50, különböző magasságban elhelyezkedő vízgyűjtő alapján

A változókból lehetséges fajlagos változókat előállítani – tulajdonképpen az említett GDP/fő is ilyen, de ilyen az időegység alatt lehullott csapadék és a csapadékos napok száma, amely a csapadék intenzitását adja meg. Amellett, hogy így lehetséges a változók számának redukálása, a csapadékintenzitás fontos meghatározója az eróziós tömegvesztésnek, hiszen a talajnak van vízfelvévőkapacitása. Így, ha 100 mm csapadék 100 nap alatt hullik le, az kisebb felszíni lemosást eredményez, mintha 10 nap alatt hullana le. Ugyanígy az eróziót befolyásolja a lejtő meredeksége, hiszen a lefolyó víz munkavégző-képessége a gravitációs erő lejtő irányú komponensétől is függ (feltételezve, hogy egyforma a mintaterületek felszínérdessége, talajszerkezete növényborítottsága, klímája, lejtőhossza, stb., mert ezek is mind befolyásoló tényezők). Így tehát négy változónk van, amelyet lehet ugyan 4/2 kétváltozós diagramon ábrázolni, de ez időigényes

és átláthatatlan. Ugyanakkor lehetséges hányadosok képzésével a változók számát kettőre (háromra) redukálni – az időegységre jutó csapadék és az egységnyi lejtőmeredekségre jutó erózió lesznek a változók (értelme lesz az egységnyi csapadékhoz tartozó erózióknak is, ez lehet a harmadik változó).

Ezek segítségével csoportképzési vagy trendvizsgálati célból mintaterületeink ábrázolhatók egy Descartes-féle koordináta rendszerben (ahol megint csak fiktív térről van szó). Az itt bemutatott fiktív adatokon alapuló példában az erózió mértékét a csapadék/nap · lejtőmeredekség képlet közelíti.

csapadék (mm)	esős napok	erózió (mm)	lejtés (%)	csapadék/esős napok	erózió/lejtés
46	24	12	5	1,92	2,40
51	26	24	10	1,96	2,40
87	12	35	7	7,25	5,00
78	10	13	2	7,80	6,50
12	6	8	4	2,00	2,00
24	12	40	24	2,00	1,67
25	8	60	24	3,13	2,50
102	20	33	5	5,10	6,60
50	11	30	6	4,55	5,00
66	30	9	8	2,20	1,13
78	11	66	15	7,09	4,40



Az adatok fiktívek!

## Nemparaméteres eljárások <sup>2</sup>

### Egymintás nemparaméteres eljárások:

Formailag ide illeszthetők az egymintás eset módszerei is, de ezen belül az egyes eljárások sokfélék:

- |                              |                       |                       |
|------------------------------|-----------------------|-----------------------|
| 1. Chi-square                | (Khi - négyzet próba) | nominális, ordinális  |
| 2. Binomial                  | (Binomiális próba)    | kétértékű (nominális) |
| 3. Runs (Wald-Wolfowitz)     | (Futamok vizsgálata)  | kétértékű (nominális) |
| 4. Kolmogorov-Szmirnov próba |                       | intervallum           |

<sup>2</sup> A leírásnál Máth János óravázlatát követjük, a feladatokat földrajzos szemléletűre cseréltük. ([http://psycho.unideb.hu/munkatarsak/math\\_janos/statisztika\\_II/nempar.doc](http://psycho.unideb.hu/munkatarsak/math_janos/statisztika_II/nempar.doc))

### 1. Khí-négyzet próba:

Adott egynelemű minta, k lehetséges értékkel. Az értékek gyakoriságai:  $n_1, n_2, \dots, n_k$ . Azt kérdezzük, hihető-e, hogy a populáció szintjén ezek a gyakoriságok  $p_1, p_2, \dots, p_k$  ?

A próba-statisztika:

$$\sum_{i=1}^k \frac{(n_i - N \cdot p_i)^2}{N \cdot p_i} = \chi_{n-1}^2$$

Adott pl. a statisztika a városok által elnyert pályázatok számáról (itt most nem az összeget, hanem a darabszámot vizsgáljuk!). Az adatbázisban minden egyes esetet fel kell tüntetni sorként (nemcsak az összesítést!), az adott városhoz ugyanazt a számot kell rendelni (pl. Eger = 1)

$H_0$ : A populációban az arányok  $p_1, p_2, \dots, p_k$ .

$H_1$ : nem  $H_0$

**Esetünkben azt vizsgáljuk, van-e a városok sikeressége között különbség az elnyert pályázatok alapján.** Ekkor  $p_1 = p_2 = \dots = p_k = 1/k$ .

Analyze - Nonparametric test – Chi-square

Az eljárás lefuttatása után az alábbi eredményeket kapjuk:

	VAR00001	VAR00002	var	var	var	var	var	var	var
1	Eger	1,00							
2	Eger	1,00							
3	Eger	1,00							
4	Eger	1,00							
5	Eger	1,00							
6	Miskolc	3,00							
7	Miskolc	3,00							
8	Miskolc	3,00							
9	Ózd	2,00							
10	Ózd	2,00							
11	Szolnok	4,00							
12	Szolnok	4,00							
13	Szolnok	4,00							
14	Szolnok	4,00							
15	Szolnok	4,00							
16	Szolnok	4,00							
17	Szolnok	4,00							
18	Szolnok	4,00							
19	Szolnok	4,00							
20	Szolnok	4,00							
21	Szolnok	4,00							
22									

	Observed N	Expected N	Residual		VAR0000 2
1,00	5	5,3	-,3	Chi-Square(a)	<b>9,286</b>
2,00	2	5,3	-3,3	df	3
3,00	3	5,3	-2,3	Asymp. Sig.	,026
4,00	11	5,3	5,8		
Total	21				

0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 5,3.

A chí-négyzet értéke alapján tehát van különbség a napok között (nagy az eltérés az elvárt – expected és megfigyelt – observed esetek között).

## 2. Binomiális próba

Tegyük fel, hogy van egy populáció, ahol a munkanélküliek aránya hosszú ideje 20 %. A mai világgazdasági helyzetben felmerül a gyanú, hogy ez az arány növekedett. Hipotézisünk helytállóságát megvizsgálandó, veszünk egy mintát, ahol 1-gyel kódoljuk a munkanélkülieket:

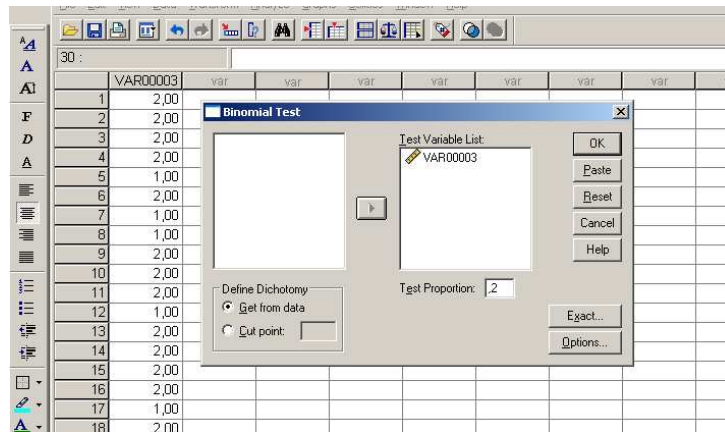
2,2,2,2,1,2,1,1,2,2,2,1,2,2,2,2,1,2,1,2,2,1,2,2,2,2,1,2,1,1

$H_0$ : A populációban a munkanélküliek aránya 20%

$H_1$ : A populációban a munkanélküliek aránya több, mint 20%

**Próbastatisztika:** a mintában a munkanélküliek száma.

**Analyze - Nonparametric test – Binomial.** A próbastatisztika  $H_0$  esetén binomiális eloszlást követ ( $n=30$ ,  $p=0.20$ ) paraméterekkel (ezt a kívánalmat a **test proportion**-nél tudjuk beállítani). Ha  $n$  nagy, a binomiális helyett normális eloszlású közelítést alkalmaznak.



### Binomial Test

	Category	N	Observed Prop.	Test Prop.	Aymp. Sig. (1-tailed)
VAR0003	Group 1	2,00	20	,7	,000(a)
	Group 2	1,00	10	,3	
	Total		30	1,0	

a Based on Z Approximation.

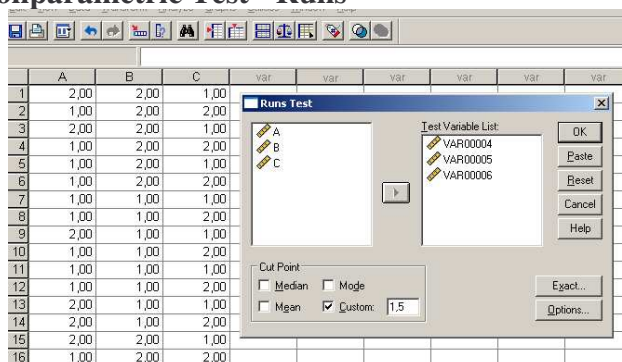
Látható, hogy a  $H_0$  szerinti gyakoriság 0.20, a tapasztalati viszont  $10/30 = 0.33$ . Az egyoldali szignifikancia - ami pont kell nekünk - 0.0175, tehát elvetjük  $H_0$ -t.

### 3. Wald-Wolfowitz féle run-próba

Fontos, hogy a mintaelemek függetlenek legyenek egymástól, mert a kapott szignifikancia értékek enélkül használhatatlanok. A Wald-Wolfowitz féle run-próba azt vizsgálja, hogy az adott kétértékű minta értékei „véletlenszerűen” követik-e egymást, azaz a fenti függetlenség fennáll-e. Mindez az ún. “futamok” számán múlik. Egy futamon az azonos értékek leghosszabb sorozatát értjük. A B változó egy 6 elemű futammal kezdődik és egy 8 elemű futammal folytatódik. Ha a minta véletlenszerű, akkor a futamok száma általában se nem túl sok, se nem túl kevés. Az alábbi példában három minta látható: az első véletlennek tekinthető, míg a két utolsó nem. Lent a hozzájuk tartozó SPSS output látható. A futamok száma az egyes mintákban **8, 3, 16**, az utóbbi két esetben a futamok száma túl kicsi, illetve túl nagy.

Az eljárás folytonos változókra is alkalmazható, ha megadunk egy értéket, amely kettévágja a lehetséges értékek tartományát (**Cut point**). Ekkor az 1-es az érték alattiakat jelenti, a 2-es az érték felettiakat. Formailag mindig ezt kell tenni. Itt az 1 és 2 között az elválasztó érték **1.5 (Custom)**.

#### Analyze -Nonparametric Test - Runs



	A	B	C
Test Value(a)	1,5000	1,5000	1,5000
Total Cases	16	16	16
Number of Runs	8	3	16
Chi-Square	,000	-2,847	3,364
Asymp. Sig. (2-tailed)	1,000	,004	,001

a User-specified.

#### 4. Kolmogorov-Szmirnov próba

Az egymintás Kolmogorov-Szmirnov próba azt vizsgálja, hogy a minta vajon adott eloszlású-e? Az SPSS normális eloszlásra, Poisson eloszlásra, és egyenletes eloszlásra (uniform) tudja tesztelni adatsorunkat. Ha a normális eloszlás paramétereit a mintából becsüljük, nem tudjuk előre, akkor a kérdés így hangzik: **az adott minta normális eloszlásból származik-e?**

Példánkban egy országos mintavételezés eredményét mutatjuk be (közel 700 eset). A vett minta a súlyos szívelégtelenséggel küzdőkre fókuszált.

#### Analyze – Nonparametric tests – One Sample K-S

		kor
N		678
Mean		66,70
Std. Deviation		11,240
Normal Parameters(a)	Positive	,043
	Negative	-,044
Kolmogorov-Smirnov Z		1,152
Asymp. Sig. (2-tailed)		,141

a Test distribution is Normal.

Az eredményekből látható, hogy bár az átlagéletkor magas (magasabb, mint a normál populációé), a szignifikancia-szint alapján közelít a normáloszláshoz ( $p > 0,05$ , legalábbis magasabb az érték, mit a többi eloszlás tesztelése esetén). A viszonylag kis szignifikancia szint oka itt az, hogy a szórás kicsi és fiatalok szinte egyáltalán szerepelnek a tesztelt adatsorban.

Ha ezek után az a kérdésünk, hogy életkoruk alapján kimutatható-e eltérés az átlagos magyar populációtól, akkor a már megismert **két mintás t-próbát** kell alkalmazni, ami viszont parametrikus teszt (**Analyze - Compare means**) (a t-próbához arányskála jellegű adatokra van szükség). Az előregedő társadalmakra (így Magyarországra) általában igaz, hogy a lakosság normáloszláshoz hasonló eloszlást mutat, hiszen viszonylag kevés a gyerek. Ehhez elegendő megvizsgálni

a korlát (korstruktúrát). A két populáció tehát összevethető (normáloszláshoz közelít mindkét eloszlás).

### Többváltozós nemparaméteres eljárások

Az eddigiek során talákoztunk már olyan eljárásokkal, melyek a változók középértékét vizsgálták: egymintás-, páros-, független mintás t-próba stb. Ezek jól használható eljárások, de alkalmazásuknak vannak feltételei, melyek nem mindig teljesülnek, pl. normalitás, azonos szórások. Ezek **paraméteres eljárások** voltak abban az értelemben, hogy feltételeztünk egy eloszlást és a vizsgálódás az eloszlás konkrét paraméterére - átlag, szórás - vonatkozott.

	paraméteres eljárások	nemparaméteres eljárások
egy minta	egymintás t-próba	KOLMOGOROV-SMIRNOV
két független minta	független mintás (kétmintás) t-próba	MANN-WHITNEY, KOLMOGOROV-SMIRNOV
két összetartozó minta	páros t-próba	WILCOXON
több független minta egy szempont szerint	egyszempontos varianci- analízis	KRUSKAL-WALLIS

A **nemparaméteres eljárások** nem kívánnak különösebb feltételt a változók eloszlásáról, sőt még az intervallum-skála sem követelmény. Az azonban jó, ha a változó folytonos vagy az intervallum finom beosztású (szűk, és az intervallumok száma nagy). Ugyanis a medián itt fontos szerepet játszik és ekkor van igazi jelentése.

A **nemparaméteres eljárások nem a tényleges értékekkel dolgoznak, hanem az értékek sorrendjével, amit az ún. rangokkal** fejeznek ki. Egy érték rangja azt jelenti, hogy ő nagyság szerint hányadik a mintában. Nézzünk erre egy példát, ahol egy változó mellett a rangok oszlopa látható.

	GDP éves növekedése (%)	RANG
Irán	5.00	4.500
USA	4.00	3.000
Ma- gyar- ország	1.00	1.000
India	8.00	6.000
Kína	9.00	7.000
EU	2.00	2.000
Oroszo.	5.00	4.500

A hét szám közül az **1** rangja **1**, mert az a legkisebb, a **9** rangja **7**, mert az a legnagyobb (elvben akár fordított skálázást is lehetne választani). Az **5** kétszer fordul elő, a **4.** és az **5.** helyen, ezért rangja ezek átlaga, azaz  $(4+5)/2=4.5$ .

Ezek után nézzük az egyes eljárásokat. A jobb áttekinthetőség kedvéért állítsuk őket párhuzamba a paraméteres eljárásokkal:

## 1. Két független minta

### Mann-Whitney próba

	csop	haszon	ranghasz
1	1,00	8,00	6,00
2	1,00	2,00	2,00
3	1,00	5,00	5,00
4	1,00	9,00	7,00
5	2,00	1,00	1,00
6	2,00	3,00	3,00
7	2,00	4,00	4,00
8			
9			
10			
11			
12			

The image shows a screenshot of the SPSS software interface. On the left is a data table with columns 'csop', 'haszon', and 'ranghasz'. Overlaid on the table is the 'Two-Independent-Samples Tests' dialog box. In this dialog, 'ranghasz' is selected as the test variable, and 'haszon' is also listed. The grouping variable is set to 'csopord[?]'. Under 'Test Type', the 'Mann-Whitney U' option is checked. In the foreground, a smaller dialog box titled 'Two Independent Samples: Define...' is open, showing 'Group 1' as 1 and 'Group 2' as 2.

Adott két településcsoport. Az egyik gazdálkodását központilag ellenőr-zik, a másikat nem (csopord oszlop). Azt vizsgáljuk, az ellenőrzés milyen hatással van a település gazdálkodására (haszon oszlop). A két csoportot összehasonlítva azt kérdezzük, az **egyik csoport tagjai nagyobb haszonkulcsot produkálnak, mint a másik csoportbeliek?** Az első csoportban 4 település van (KOD=1), a másikban 3 (KOD=2). Utána a két mintát egyesítve kiszámoljuk a rangokat, majd az egyes csoportokba eső rangokat átlagoljuk. Ez az érték az első csoportban  $(6+2+5+7)/4=5$ , a másikban pedig  $(1+3+4)/3=2,67$ . Ez látható lent. Ha a két el-oszlás egybeesik, tehát a központi kontrollnak nincs hatása, akkor e két rangátlag közel lesz egymáshoz. Ezen alapszik a próbastatisztika, melyhez tartozó szignifikancia érték alapján dönthetünk a hipotézisekről, melyek az alábbiak:

$H_0$  : A két csoport értékei között nincs szisztematikus eltérés

$H_1$  :  $H_0$  nem igaz.

**Analyze – Nonparametric test – 2 Independent Samples:** beállítjuk a vizsgált változót (haszon) és a csoportosítási kritériumot (**Grouping variable**), ahol meg kell adni a csoportokat elkülönítő értékeket (itt 1 és 2). Majd kiválasztjuk alul a megfelelő próbát.

Ha a két csoport eloszlása - átlagtól eltekintve - megegyezik, akkor a fenti próba az átlagok egyenlőségét teszteli, mint a kétmintás t-próba.



Ranks

	CSOPORT	N	Mean Rank	Sum of Ranks
HASZON	1,00	4	5,00	20,00
	2,00	3	2,67	8,00
	Total	7		

Test Statistics	HASZON
Mann-Whitney U	2,000
Wilcoxon W	8,000
Z	-1,414
Asymp. Sig. (2-tailed)	<b>,157</b>
Exact Sig. [2*(1-tailed Sig.)]	,229

a Not corrected for ties.

b Grouping Variable: CSOPORT

A két csoport átlagos rangja (5 és 2,68) között nem erősen szignifikáns a különbség ( $p=0,157$ ).

Ha az eloszlások - különös tekintettel a szélsőséges esetekre - nem tekinthetők normálisnak, a t-próba valóban nem a legjobb választás.

**Kolmogorov-Szmirnov próba**

Két független minta összevetésének egy másik módja az ún. Kolmogorov-Szmirnov próba. Ez az eljárás a két minta tapasztalati eloszlásfüggvényét veti össze és a tapasztalt eltérésekből készít próbastatisztikát. Hipotézisei:

**H<sub>0</sub>**: a két eloszlás megegyezik

**H<sub>1</sub>**: a két eloszlás nem egyezik meg

Az eloszlások közötti bármilyen jellegű eltérésre érzékeny.

Most a munkanélküliek arányát (%) vizsgáljuk két kistérségben, az egyikbe 5, a másikba négy elem tartozik. Itt is használjuk a **Grouping variable, Define Group** parancsot a kategóriák meghatározására és válasszuk ki a megfelelő próbát.



Test Statistics		MUNKANEL
Most Extreme Differences	Absolute	1,000
	Positive	1,000
	Negative	,000
Kolmogorov-Smirnov Z		1,491
Asymp. Sig. (2-tailed)		<b>,023</b>

a Grouping Variable: REGIO

A szignifikancia szint alapján ( $p=0,023$ ) a nullhipotézist elvetjük, a két eloszlás nem egyezik meg (ami nem meglepő, hiszen az egyik csoportban folyamatosan növekvő értékek szerepelnek, de csak egy-egy érték tartozik ide, a másikban pedig ismétlődő értékek dominálnak).

## 2. Két összetartozó minta

### Wilcoxon-próba

A Wilcoxon próba közeli rokona a Mann-Whitney próbának. Ugyanúgy a rangokkal dolgozik. Hipotézisei:

$H_0$ : a két minta(sor) különbségének mediánja **nulla**

$H_1$ : a két minta(sor) különbségének mediánja **nem nulla**

A próba menetét ismét egy kicsi adatbázison mutatjuk meg. Adott **A** és **B** változó, 7 választópolgár véleménye egy politikusról választási beszéde és után (10-es skálán értékelték). Előbb kiszámítjuk a két minta páronkénti különbségét (C oszlop), majd abszolút értékét vesszük (ABS\_C) és ezen oszlop rangjait számítjuk ki.

A	B	c	ABS_C	
RANGABS_C				
5.00	6.00	-1.00	1.00	1.500
4.00	7.00	-3.00	3.00	5.000
1.00	2.00	-1.00	1.00	1.500
<b>8.00</b>	<b>5.00</b>	<b>3.00</b>	<b>3.00</b>	<b>5.000</b>
<b>9.00</b>	<b>7.00</b>	<b>2.00</b>	<b>2.00</b>	<b>3.000</b>
2.00	5.00	-3.00	3.00	5.000
5.00	9.00	-4.00	4.00	7.000

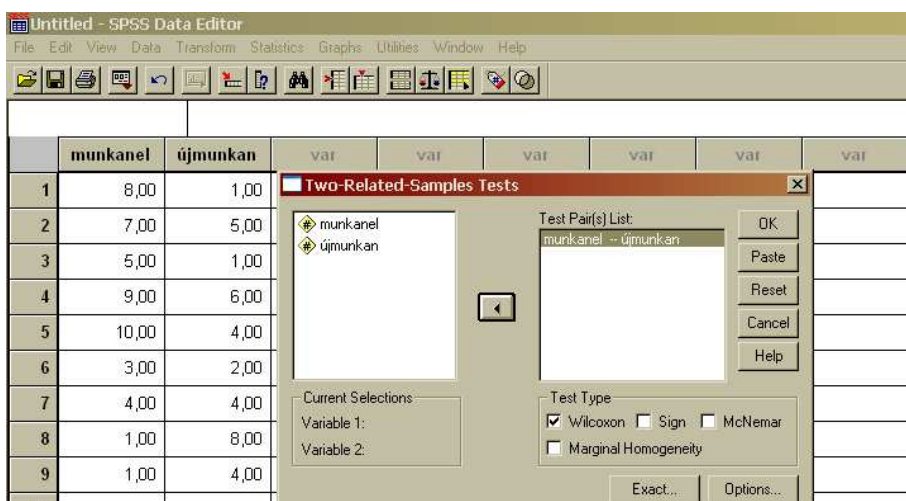
Utána külön tekintjük azon rangok átlagát, ahol a különbség pozitív volt és azt is, ahol negatív. Az első csoportban két rang van: 5 és 3. Ezek átlaga 4. A másik csoportban ez a szám:  $(1.5+5+1.5+5+7)/5=4$ . Ezen adatokból számíthatjuk ki a próbastatisztika értékét, és a hozzá tartozó szignifikancia szintet. **Ha a két átlag-rang közel van egymáshoz, akkor a  $H_0$  lesz igaz.** Itt a szignifikancia szint  $0,3105$

( $p>0,05$ ), tehát valóban a  $H_0$  mellett maradunk.

----- Wilcoxon Matched-Pairs Signed-Ranks Test	
A with B	
Mean Rank	Cases
<b>4.00</b>	2 - Ranks (B LT A)
<b>4.00</b>	5 + Ranks (B GT A)
	0 Ties (B EQ A)
	-
	7 Total
Z = -1.0142	2-Tailed p = <b>.3105</b>

Tekintsünk egy régiót, ahol a munkanélküliség értékeit látjuk (%-ban) 1990-ben és 2000-ben. Vizsgáljuk meg, hogy a rangokat felhasználva, hogy a két átlag rang közel esik-e egymáshoz?

## Analyze – Nonparametric tests – 2 Related samples



Ranks		N	Mean Rank	Sum of Ranks		
ÚJMUNKAN - MUNKANEL	Negative Ranks	6	4,17	25,00	Z	-0,93
	Positive Ranks	2	5,50	11,00	Asymp. Sig.	<b>0,326</b>
	Ties	1				

- a ÚJMUNKAN < MUNKANEL
- b ÚJMUNKAN > MUNKANEL
- c MUNKANEL = ÚJMUNKAN

A szignifikancia szint ( $p=0,326$ ,  $p>0,05$ ) és a két átlag alapján nem érzékelhető jelentős különbség a rangokat (és nem a változók tényleges értékeit, hiszen az egy két mintás t-próba lenne!) összevetve, a nullhipotézis érvényes.

	1960	1970
New York	53756	43567
Washington	23891	23456
Denver	25612	25110
Chicago	69567	35678
New Orleans	53987	66123
Los Angeles	36452	23567
San Francisco	21345	11234

### Feladat:

Bűnözésföldrajzi példa is megoldható a módszerrel: pl. ismeretesek az amerikai városokban történt bűnesetek (páronként összetartoznak) 1960-ban és 1970-ben. Van-e szignifikáns eltérés a két adatsor között, ha a települések rangját használjuk fel a vizsgálatban?

### 3. Több független minta

#### Kruskal-Wallis próba

Ez az eljárás a Mann-Whitney próba általánosítása, ekkor nem kettő, hanem több független mintánk van. Kérdés, van-e **a több** minta között szisztematikus eltérés?

A számolás logikájában is nagyon hasonlít a Mann-Whitney-re. Egyesíti a három (vagy több) mintát és kiszámítja a rangokat. Ezután a három csoport rangjait külön átlagolja, majd ebből számítja a próbastatisztikát. Ha a három átlag-rang közel van egymáshoz, akkor nincs eltérés az adatsorok között.

Adott két kistérség (var0002) és három időpontból a munkanélküliség (%). Ismétcsak nemparaméteres tesztet hajtunk végre. **Analyze – Nonparametric test – K Independent Samples.** Megint definiáljuk a csoportokat. Az eredményekből látható, hogy nincs nagy különbség a rangok átlagai között

Ranks	VAR002	N	Mean Rank	Test Statistics	MUNKANEL	ÚJMUNKAN	MÉGÚJABB
MUNKANEL	1,00	5	4,20	Chi-Square	,968	,063	,016
	2,00	4	6,00	Asymp. Sig.	<b>,325</b>	<b>,802</b>	<b>,900</b>
ÚJMUNKAN	1,00	5	4,80	a Kruskal Wallis Test b Grouping Variable: VAR00002			
	2,00	4	5,25				
MÉGÚJABB	1,00	5	4,90				
	2,00	4	5,13				

## *Nominális változók*

### **Függetlenség vizsgálat**

Ha a két változó nominális vagy ordinális, a függetlenség-vizsgálat a már említett Khi-négyzet próbával lehetséges.

### **Homogenitás vizsgálat**

Formailag ugyanúgy történik, mint a függetlenség-vizsgálat, csak más az értelmezése. Mindkét esetben azt kérdezzük, az egyik változó eloszlása eltérő-e a másik változó különböző értékeinél. Ha a függetlenség vizsgálat során azt kapjuk, hogy a két változó független egymástól, akkor a kérdést le is zárhatjuk. Ha azonban nem függetlenek, akkor a kapcsolat mibenlétét, erősségét kezdhetjük firtatni. Erre szolgálnak a **különböző asszociációs mérőszámok**.

Az eredményeket az SPSS-ben a **Crosstabs** parancs szolgáltatja.

### **Nominális változók kapcsolata**

Tekintsük az alábbi példát, mely a bűnözés és az ital kapcsolatát vizsgálja. Két változónk van:

- A bűncselekmény fajtája
  1. gyújtogatás
  2. garázdaság
  3. szélhámosság
- Az italhoz való viszony
  1. iszik
  2. nem iszik

**Minél nagyobb a  $\chi^2$  értéke, annál távolabb vagyunk a függetlenségtől.** A Khi-négyzet statisztikából transzformáció útján többféle mérőszámot szoktak kiszámolni (phi-együttható, kontingencia-együttható, Cramer-féle V). Ugyanis az a cél, hogy a kapott érték olyan határok közé essen, melyek jelentése egyértelmű (összevethető legyen több vizsgálat). Például a korreláció esetében az együttható -1 és 1 közé esett és e szélsőértékek jelentése is egyértelmű: a khi négyzet értékéről viszont nem tudjuk mekkora, mivel az alsó és felső határ nincs definiálva. Itt a 0 és az 1 a két vágyott végpont.

```
CROSSTABS
  /TABLES=ital BY bun
  /STATISTIC=CHISQ CC PHI LAMBDA
  /CELLS= COUNT TOTAL .
```

ITAL by BUN

Page 1 of 1

Count   Tot Pct	BUN			Row Total
	gyűjtoga ltás   1.00	garázdas ág   2.00	szélhámo sság   3.00	
ITAL	-----+-----+-----+-----+			
1.00	50	155	63	268
iszik	8.8	27.4	11.2	47.4
	+-----+-----+-----+			
2.00	43	110	144	297
nem iszik	7.6	19.5	25.5	52.6
	+-----+-----+-----+			
Column	93	265	207	565
Total	16.5	46.9	36.6	100.0

Chi-Square	Value	DF	Significance
-----	-----	----	-----
<b>Pearson</b>	<b>38.47692</b>	<b>2</b>	<b>.00000</b>
Likelihood Ratio	39.27545	2	.00000
Mantel-Haenszel test for linear association	24.38375	1	.00000

Minimum Expected Frequency - 44.113

Statistic	Value	ASE1	Approximate Val/ASE0	Significance
-----	-----	-----	-----	-----
<b>Phi</b>	<b>.26096</b>		<b>.00000 *1</b>	
<b>Cramer's V</b>	<b>.26096</b>		<b>.00000 *1</b>	

**Contingency Coefficient**      **.25250**      **.00000 \*1**

Lambda :				
symmetric		.15141	.04797	3.00525
with ITAL	dependent	.19403	.06338	2.76684
with BUN	dependent	.11333	.05002	2.14199
Goodman & Kruskal Tau :				
with ITAL	dependent	.06810	.02078	.00000 *2
with BUN	dependent	.04328	.01361	.00000 *2

### Phi együttható

$$\Phi = \sqrt{\frac{\chi^2}{N}} = \sqrt{\frac{38,48}{565}} = 0,26$$

A Phi együttható értéke lehet 1-nél nagyobb is, ha táblázat sorainak vagy oszlopainak száma meghaladja a 2-t, e miatt nem igazán jó mérőszám. A túlsó véglet, **a 0 érték akkor áll elő, ha a Khí-négyzet értéke 0, vagyis függetlenség esetén.** Jelen esetben tehát nem beszélhetünk a két adatsor (nominális változó) függetlenségéről.

Egy másik, Khi-négyzet alapú statisztika az un.

### Kontingencia együttható

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}} = \sqrt{\frac{38,48}{38,48 + 565}} = 0,25$$

A kontingencia együtthatóval az a probléma, hogy bár 0 és 1 között marad, sosem éri el az 1-et. A 0 érték most is ugyanazt jelenti.

A harmadik Khí-négyzet alapú statisztika az ún.

### Cramer féle V együttható

$$V = \sqrt{\frac{\chi^2}{N(k-1)}} = \sqrt{\frac{38,48}{565}} = 0,26$$

ahol  $k$  az oszlopok és sorok száma közül a kisebb. Ez a mérőszám már elérheti az 1-et is bármely táblázat esetén. Ha két oszlopunk vagy sorunk van, mint most is, akkor értéke azonos a **Phi** együtthatóval.

A szignifikancia szint mindhárom esetben ugyanazt jelenti:

$H_0$ : a két változó független

$H_1$ : a két változó nem független

### Ordinális változók

Az alábbi példában gyermekek mentális állapotának és a szülők társadalmi helyzetének kapcsolatát vizsgáljuk. Ugyanúgy a khí-négyzet értéke dönti el, hogy független-e egymástól a két változó (a szülők társadalmi helyzete és a gyermek mentális helyzete):

```
CROSSTABS
/TABLES=gyerek BY szulo
/STATISTIC=CHISQ GAMMA D BTAU CTAU CORR
/CELLS= COUNT TOTAL .
```

SPSS Output

GYEREK gyerek mentális helyzete by SZULO szülő társadalmi pozíciója					
Page 1 of 1					
Count	Tot Pct	SZULO			Row Total
		1.00 alacsony	2.00 közepes	3.00 magas	
121	7.3	129	57	307	18.5
188	11.3	246	168	602	36.3
198	11.9	296	257	751	45.2
507	30.5	671	482	1660	100.0

Chi-Square	Value	DF	Significance
Pearson	31.51284	4	.00000
Likelihood Ratio	32.38933	4	.00000
Mantel-Haenszel test for linear association	30.04890	1	.00000

Minimum Expected Frequency - 89.141

### 13.2. Korrelációs számítás

A korrelációs számítás során két normális eloszlású változó közötti összefüggés szorosságát vizsgáljuk. Legyen két mérési sorozatunk a következő mérési eredményekkel:  $x (x_1, x_2, x_3, \dots, x_n)$  és  $y (y_1, y_2, y_3, \dots, y_n)$

Normális eloszlású változók között elméletileg csak lineáris kapcsolat lehetséges. Ha a koordináta-rendszerben ábrázolt pontok görbe vonalú összefüggésre utalnak, akkor legalább az egyik változó eloszlása eltér a normálistól. A korrelációs együttható ( $r$ ) a két minta közötti kapcsolat szorosságát mutatja. Az együttható értéke  $-1$  és  $+1$  közötti racionális szám lehet. Amennyiben a két minta normális eloszlásánál a korrelációs együttható értéke nulla, akkor a két változó között nincs lineáris összefüggés. Ha  $r = +1$ , vagy  $r = -1$ , akkor a változók között lineáris függvénykapcsolat van. Ha a két változó értékeit koordináta-rendszerben ábrázoljuk, akkor a pontok mind egy egyenesen helyezkednek el. A természeti és



társadalmi folyamatokra több tényező hat, ezért a  $\pm 1$  értékkel ritkán találkozunk. Általában azt szoktuk mondani, hogy ha az  $|r| = 0 - 0,3$  közötti, akkor gyenge,  $|r| > 0,6$ , akkor közepes és  $|r| > 0,8$  értékeknél már szoros az összefüggés. A pozitív korreláció azt jelenti, hogy nagyobb  $x$  értékhez várhatóan nagyobb  $y$  tartozik. Ha az egyik változó értékeinek növekedésével a másikhoz tartozók általában csökkennek, akkor a korrelációs együttható értéke negatív.

A korrelációs kapcsolat szorosságát különféle mérőszámokkal határozhatjuk meg. A lineáris típusú korrelációs kapcsolat mérőszámai: a **kovariancia** ( $C$ ), a **korrelációs együttható** ( $r$ ), és a **determinációs együttható** ( $r^2$ ).

A **kovariancia mutató** ( $C$ ) számítására a következő képletet használjuk:

$$C = \frac{1}{n} \cdot \sum [(x - \bar{x}) \cdot (y - \bar{y})] = \frac{\sum d_x \cdot d_y}{n} = \frac{1}{n} \cdot [(\sum x \cdot y) - n \cdot \bar{x} \cdot \bar{y}] = \frac{\sum x \cdot y}{n} - \bar{x} \cdot \bar{y}$$

ahol  $d_x = (x - \bar{x})$  és  $d_y = (y - \bar{y})$ ,  $\bar{x}$  az első minta számtani közepe:

$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ ,  $y_i$  a második minta aktuális eleme és  $\bar{y}$  a második minta számtani

közepe:  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$

A kovariancia tulajdonságai a következők:

- A mutató előjele a kapcsolat irányát mutatja.
- A kovariancia mutató értéke nulla, ha a pozitív és negatív előjelű eltérésszorozatok kiegyenlítik egymást, tehát az  $x$  és  $y$  változók korrelálatlanok.
- A mutató abszolút mértékének nincs határozott felső korlátja.
- Ha az  $x$  önmagával való kapcsolatát vizsgálánk, akkor  $d_x d_y$  helyébe  $d_x^2$  kerül, tehát a mutató egyenlő az  $x$  szórásnégyzetével.
- A mutató a két változóban szimmetrikus,  $x$  és  $y$  szerepe a formulában felcserélhető.

A kovariancia mutatóval csak a kapcsolat irányát lehet jelezni. A kovariancia értéke függ a változók mértékegységétől, az eltérések nagyságrendjétől és a kapcsolat szorosságától. *A kovariancia akkor maximális, ha a két adat között függvényyszerű a kapcsolat, ekkor a mutató értéke a változók szórássainak szorzatával egyenlő.*

Ha összehasonlítjuk a kiszámított kovariancia értékét, az aktuális változók feltételezett maximális kovariancia értékével, akkor olyan értéket kapunk, amely

előjelében a kapcsolat irányát mutatja, abszolút értékében a kapcsolat nagyságát is kifejezi. Az így számítható mutató a korrelációs együttható.

A **korrelációs együttható** ( $r$ ) a következő képlettel határozható meg:

$$r = \frac{C}{C_{\max}} = \frac{C}{\sigma_X \cdot \sigma_Y} = \frac{\sum d_x \cdot d_y}{\sqrt{d_x^2 \cdot d_y^2}} = \frac{\sum [(x - \bar{x}) \cdot (y - \bar{y})]}{\sqrt{(x - \bar{x})^2} \cdot \sqrt{(y - \bar{y})^2}} =$$

$$= \frac{[\sum (x \cdot y)] - n \cdot (\bar{x} \cdot \bar{y})}{\sqrt{[\sum (x^2)] - n \cdot (\bar{x}^2)} \cdot \sqrt{[\sum (y^2)] - n \cdot (\bar{y}^2)}}$$

Ahol  $r$  a korrelációs együttható, az  $x_i$  az első minta aktuális eleme,  $\bar{x}$  az első minta számtani közepe:  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ ,  $y_i$  a második minta aktuális eleme és

$\bar{y}$  a második minta számtani közepe:  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$

A korrelációs együttható tulajdonságai:

- A mutató előjele a kapcsolat irányát mutatja.
  - ha  $-1 < r < 0$ , akkor negatív irányú kapcsolatról beszélünk.
  - ha  $0 < r < +1$ , akkor pozitív irányú a kapcsolat.
- A mutató abszolút értéke a kapcsolat erősségét fejezi ki.
  - ha  $|r| = 1$ , akkor függvényszerű kapcsolat van a változók között.
  - ha  $|r| = 0$ , akkor korrelálatlanság jellemzi a két változó közti kapcsolatot.
  - ha  $0 < |r| < 1$ , akkor sztochasztikus kapcsolat van a két változó között.
    - Minél közelebb van a mutató értéke a nullához, annál gyengébb; minél inkább az egyhez tart, annál erősebb közöttük a sztochasztikus kapcsolat.
- Az X és Y szerepe felcserélhető.

A **determinációs együtthatót** ( $r^2$ ) a korrelációs együtthatóból kaphatjuk négyzetre emeléssel. Miután a négyzetre emeléssel csak pozitív értéket kapunk, így ez a mutató már "csak" a kapcsolat erősségét mutatja. A determinációs együttható számítása:  $r^2$

A determinációs együttható tulajdonságai:

- A mutató a kapcsolat erősségét határozza meg %-os formában.

ha a mutató értéke 0, az azt jelenti, hogy a változók között nincs kapcsolat.

ha a mutató értéke 100 %, az azt jelenti, hogy a változók között függvény-szerű a kapcsolat.

ha a mutató értéke 0 és 100 % közé esik, az azt jelenti, hogy az X változó hány %-ban magyarázza az Y változó átlagtól való eltérésnégyzetének átlagát.

- Az X és az Y szerepe felcserélhető a számítás során.

Ha az eredeti adatokból kivonjuk az átlagot, és elosztjuk a szórással, akkor azt mondjuk, hogy **standardizáljuk** az adatsorokat. Az így kapott adatsorok legyenek:  $x_1^*, x_2^* \dots x_n^*$ , illetve  $y_1^*, y_2^* \dots y_n^*$ . A standardizálásra azért van szükség, hogy a mérőszám ne függjön sem a mértékegység megválasztásától, sem a skála kezdőpontjától (pl. Celsius– helyett Fahrenheit–fok) nem változtatna a korrelációs együtthatón.

Az r korrelációs együttható az összetartozó standardizált adatok felhasználásával a következő képlettel számolható:

$$r = \frac{1}{n} \sum_{i=1}^n x_i^* \cdot y_i^*$$

Amint fentebb már láttuk a korrelációs együttható értéke mindig -1 és +1 közé esik, és minél jobban eltér a 0-tól, annál nagyobb az összefüggés a két adatsor között. Az előjele pedig az összefüggés irányát mutatja.

Könnyen megadhatók olyan adatok is, amelyekben az  $x_i$  és az  $y_i$  között függvénykapcsolat van, mégis 0 a korrelációs együttható. Ez azt mutatja, hogy kis abszolút értékű korrelációból még nem feltétlenül következtethetünk mindenfajta összefüggés hiányára. Ám ez sem magyarázza a korrelációs együttható negatív előjeléből adódó fordított irányú összefüggést.

A háttérváltozó hatásának számszerű kiküszöbölésére a **parciális korrelációs** együtthatót használják. Ezt úgy kapjuk meg, hogy mindkét adatsorból levonjuk a „c” háttérváltozónak azt a lineáris kifejezését, amely a legjobban közelíti a négyzetes hibában. Megkeressük tehát azokat az együtthatókat, amelyekre a négyzetes hibák összege a legkisebb, ezt a matematikai statisztikában lineáris regressziós feladatnak nevezik, majd képezzük az úgynevezett reziduálisokat.

Ezek az x változó olyan értékei, amelyekből leszámítjuk a „c” háttérváltozó hatását. Hasonló módon képezzük az  $y_i$  korrigált értékeket is, végül a két módosított adatsornak számítjuk ki a korrelációs együtthatóját.

A végeredmény kifejezhető az adatsorok eredeti korrelációs együtthatóival is:

$$r_{xy \cdot c} = \frac{r_{xy} - r_{xc} \cdot r_{yc}}{\sqrt{(1 - r_{xc}^2) \cdot (1 - r_{yc}^2)}}$$

Ha kettőnél több változó kapcsolatrendszerét vizsgáljuk, akkor *lineáris többszörös korrelációról* beszélünk és magát az együtthatót **R**-el jelöljük. Az **R** értéke 0 és 1 között van: 0 érték esetén a változók között nincs lineáris kapcsolat, míg 1 esetén a kapcsolat maximális erejű. Az  $R^2$  determinációs együtthatót a két-változós  $r^2$ -hez hasonlóan értelmezzük:

$$R^2 = \frac{s_{1 \cdot 23}^2}{s_1^2}$$

ahol  $s_{1 \cdot 23}^2$  az  $x_1$  varianciájának az a része, amit az  $x_2$  és  $x_3$  változók együttesen magyaráznak, valamint  $s_1^2$  az  $x_1$  változó teljes varianciája.

A parciális korrelációnál két változó kapcsolatát úgy vizsgáljuk, hogy a többi változó hatását konstansnak tekintjük.

Pl. legyen három változónk  $x_1$ ,  $x_2$  és  $x_3$ , a közöttük lévő korrelációk  $r_{12}$ ,  $r_{13}$  és  $r_{23}$ . Az  $r_{12}$  hatásából az  $x_3$  hatását a következő módon szűrjük ki (elsőrendű parciális korrelációs együttható)

$$r_{123} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{(1 - r_{13}^2) \cdot (1 - r_{23}^2)}}$$

Az  $r_{123}$  együtthatót *parciális együtthatónak* nevezzük: az 1,2,3 indexben a vessző utáni szám jelenti azt a változót, amelynek hatását kiszűrjük. Az  $r_{123}$  a reziduálok közötti korrelációt jelenti, az  $x_3$  hatásának kiszűrése után.

A parciális korrelációt több tényező kiszűrésére is ki lehet terjeszteni. A parciális korrelációs együttható szignifikanciáját, a  $H_0: r_{12,3} = 0$  hipotézist, a következő statisztikával ellenőrizhetjük

$$t = \frac{r_{123}}{\sqrt{\frac{1 - r_{123}^2}{N - 3}}}$$

amely  $f = n - 3$  szabadságfokú t-eloszlást követ.

### 13.3. Regressziószámítás

A regressziószámítás a változók egymásra gyakorolt hatásának számszerűsítésével, e hatások irányának és mértékének megállapításával foglalkozik. A

regressziószámítás feltétele, hogy az  $y$  változó eloszlása minden  $x$  értéknél normális legyen, azonos szórással.

A legegyszerűbb a grafikus ábrázolás módszere, amely alkalmas a korreláció típusának meghatározására, leolvasható belőle a kapcsolat iránya. Ehhez a tapasztalati adatok állnak általában rendelkezésre, amelyek elemzési eszköze az ún. *tapasztalati regresszió függvény*.

Akkor, ha nem ismert pontosan a változók közötti kapcsolat, akkor *analitikus regresszió függvény* segítségével, analitikus regressziószámítással keressük a kapcsolatot.

A lineáris regressziószámításnál azt a függvényt ( $y = f(x)$ ) határozzuk meg, amelynek a segítségével adott  $x$  (független változó) értékhez az  $y$  (függő) változó legjobb becslését kaphatjuk meg. Az egyenes képletében ( $y = \mathbf{m}x + \mathbf{b}$ )  $\mathbf{m}$  a meredekség és  $\mathbf{b}$  a tengelymetszet. Az  $\mathbf{m}$  azt mutatja meg, hogy az  $x$  változó egységnyi emelkedésével várhatóan hogyan változik az  $y$ . Ezt az egyenest az jellemzi, hogy a koordináta-rendszerben elhelyezkedő pontok és az egyenes között  $y$  irányban mért távolságnégyzetek összege a lehető legkisebb (ld. legkisebb négyzetek módszere).

A számítás előtt célszerű derékszögű koordináta-rendszerben ábrázolni a pontokat, mert ez segítséget nyújthat a jelentős eltérések kimutatásában. Az ábrán már az is látszik, hogy a pontok körülbelül egyforma szélességű sávot fednek-e le, tehát a szórások egyenlőségének feltétele várhatóan teljesül, vagy egyes elemek eltérő értékei miatt a szórás eltérő. A szórások eltérése esetén transzformáció alkalmazására van szükség (pl. ha a két változó közötti összefüggés nem lineáris). A leggyakrabban alkalmazott transzformációk a logaritmusos, az exponenciális vagy a hatványozás. Ezeket szükség esetén akár a független, akár a függő, akár mindkét változóra alkalmazhatjuk. A regressziószámítás elvégzése után a következő eredményeket kapjuk:

- Az egyenes *meredekségének* becslése ( $\mathbf{m}$ ).
- A *tengelymetszet* becslése ( $\mathbf{b}$ ).
- Az egyenes körüli *megbízhatósági sáv*
- Az egyenes körüli *tolerancia- vagy predikciós sáv*

Regressziószámítási feladat megoldásakor nem a korrelációt, hanem annak négyzetét elemezzük/értékeljük. Tehát az  $y = f(x)$  regressziós függvényt a legkisebb négyzetek elve és módszere alapján határozzuk meg. Ilyenkor az a követelmény, hogy az adott függvénytípus használata során a

$$\sum_{i=1}^n (y_i - \bar{y}_i)^2$$

minimális legyen. Az  $y_i - \bar{y}_i$  eltérések (rezidiumok) négyzeteinek összege jól jellemzi a pontthalmaz és a regressziós vonal kölcsönös viszonyát.

Ez ugyanis azt adja meg, hogy hányad részével csökken a függő változó varianciája az eredetihez képest, ha az illetett egyenestől való eltérések alapján számoljuk ki.

A legkisebb négyzetek módszerével előállított egyenest, ha a függőleges távolságokra minimalizálunk, akkor első regressziós egyenesnek, ha a vízszintes távolságokra minimalizálunk, akkor pedig második regressziós egyenesnek nevezük. A lineáris korreláció analízis lényegében azt vizsgálja, hogy mennyire tér el az első regressziós egyenes meredeksége a másodiktól. A lineáris korrelációt a tapasztalati adatokból a fentebb megadott képlettel ( $r =$ ) határozhatjuk meg.

A lineáris regresszió vizsgálat során általában az első regressziós egyenest alkalmazzuk. Ekkor az  $\mathbf{m}$  és  $\mathbf{b}$  becslült értékeire a legkisebb négyzetek módszerét alkalmazva a következőt kapjuk:

$$m = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \bar{y} = m \cdot \bar{x} + b$$

ahol „ $m$ ” (iránytangens) a regressziós együttható azt mutatja, hogy  $x$  egységnyi változása mekkora változást idéz elő  $y$ -ban, és a „ $b$ ” az  $x=0$  helyhez ad regressziós becslést

A lineáris regresszió jele:  $R^2$ , értéke 0 és 1 közé esik, tehát negatív szám nem lehet. Az  $R^2$  értéke egynél több független változó esetén is megadható. Ebben az esetben  $R$  már nem jelent korrelációt. Az  $R^2$  azt jelenti, hogy a független változók együttesen mekkora részét „magyarázzák” a függő változó variabilitásának.

#### 13.4. A korrelációs együttható szignifikancia vizsgálata

$H_0: R = 0$  A két változó egymástól független normális eloszlású

Ha  $H_0$  igaz, akkor  $r$  alábbi függvénye  $f=n-2$  szabadság fokkal  $t$ -eloszlást követ:

$$t = r \cdot \sqrt{\frac{n-2}{1-r^2}}$$

Ha adott  $\mathbf{b}$  mellett  $t_{sz} > t_{krit}$ , akkor  $H_0$ -t elvetjük és  $e=1-b$  megbízhatósággal állítjuk, hogy a két változó között sztochasztikus kapcsolat áll fenn.

### 13.5. A regresszió becslés pontossága

A sztochasztikus kapcsolat mérőszámaiból csak akkor vonhatunk le helyes következtetéseket, ha nagyszámú adatsorral dolgozunk. Így, az eredmények értékeléséhez hozzátartozik a mérőszámok hibájának vizsgálata is. A pontosság jellemzése céljából tehát most az **m**, **b**, paraméterek becslésének szórását (standard hibáját) kell meghatároznunk:

- A regressziós együtthatók standard hibái (pontbecslés).
- Konfidencia intervallum a becsült paraméterekre.
- A lineáris kapcsolat szignifikancia vizsgálata.
- Az átlagos, vagy az egyedi  $y_i$  értékek becslése.

A lineáris kapcsolat szignifikancia vizsgálata  $t$ -próba segítségével azt is ellenőrizhetjük, hogy az  $y$  és  $x$  változók között szignifikáns lineáris kapcsolat van-e. Nullhipotézisünk és ellenhipotézisünk:  $H_0: m=0$  és  $H_0 \neq 0$

$$\text{a próbastatisztika: } t_{sz} = \frac{m}{s_m}$$

A  $t_{krit}$  értéket **b** szignifikancia szinten  $f=n-2$  szabadsági foknál találjuk meg. Ha  $t_{sz} > t_{krit}$ , elvetjük  $H_0$ -t és valós lineáris összefüggést tételezünk fel  $x$  és  $y$  között.

### 13.6. Idősorok elemzése

A geográfiában gyakran időrend szerint felsorolt adatokkal dolgozunk. Az idősorok adatainak elemzése a korrelációs számítás szempontjából speciális problémát vet fel. Az idősorok elemzésénél valamely adatsort egy időbeli ismérv szerint rendezzük, s ezáltal vizsgáljuk a statisztikai jelenség időbeli változását, alakulását. Az idősorok elemzésével a múlt időszak tendenciáit, összefüggéseit tárhatjuk fel, és ennek alapján a jövőre nézve is fontos információkat kaphatunk.

Gyakran előfordul, hogy egy vagy több idősor egymást követő adatai egymástól nem függetlenek, hanem szoros korrelációban állnak egymással. Ha egy adott idősor egymást követő adatainak kapcsolatát vizsgáljuk, akkor autokorrelációt végzünk. A két vagy több változó idősora közötti korrelációt kereszt-korrelációnak szokták nevezni.

Az idősorokat általában két csoportra szokták bontani. Állapot idősorról akkor beszélünk, ha adatsorunk az egy-egy időpontra vonatkozik (pl. Az egyetemi hallgatók száma a Debreceni Egyetemen – az adatok szept. 1-re vonatkoznak). Ettől eltérően a tartam idősoroknál az időbeli változását az időszakok, időtartamok egységei jelzik (pl. a Magyarországra érkező külföldi turisták számának alakulása az elmúlt 10 évben – évente).

Egy jelenség időbeli alakulását a következő módszerekkel vizsgálhatjuk:

- Az időbeli folyamatot szemléltethetjük grafikus úton (pl. diagram).
- Viszonyszámokkal kifejezhetjük az időbeli változások relatív viszonyait.
- A vizsgált időszakra meghatározhatunk egy átlagos tendenciát.
- Meghatározhatjuk, hogy az általunk meghatározott időszakban milyen átlagos változás jellemzi az adott jelenséget.
- Megvizsgálhatjuk, hogy milyen összetevők befolyásolják az adott időben lezajlott folyamatot.

Az idősor elemzés analitikus vizsgálata azt jelenti, hogy a most felsorolt szempontok vizsgálatát hajtjuk végre.

#### *Az idősorok elemzésének egyszerűbb módszerei:*

- Dinamikus viszonzszámok számításával a jelenségek időbeli változásának mértékét, vagy ütemét tudjuk meghatározni.

$$V_d = \frac{(x_1)}{(x_0)}, \text{ ahol } V_d \text{ a dinamikus viszonzszám, } x_1 \text{ a tárgyévi adat, } x_0 \text{ pedig a bázisévi adat.}$$

- Bázisviszonzszámok egy vizsgált jelenség változásának mértékét fejezi ki.  
Jele:  $b_t$  ( $t$ -edik időszakra/időpontra)

$$b_t = \frac{(x_t)}{(x_0)} \text{ ahol } x_t: \text{ a } t\text{-edik időszak/időponthoz tartozó adat és } x_0: \text{ a bázis időszak/időponthoz tartozó adat.}$$

- Grafikus ábrázolás különböző típusú ábrázolási módokkal (vonaldiagram, oszlop- és szalagdiagram, kör és egyéb diagram, térkép/kartogram, piktogram). Fontos, hogy a legkifejezőbb ábrázolási módot válasszuk, és a skála, jelmagyarázat egyértelmű legyen.
- Átlagos értékek meghatározása attól függ, hogy az adatok állapot, vagy tartam időpontra vonatkoznak. Az állapotidősor adatainak átlagolására alkalmazhatunk kronológikus átlagot ( $\bar{x}_k$ ), amelynek a képlete:

$$\bar{x}_k = \frac{x_1 + x_2 + x_3 + \dots + x_{n-1} + x_n}{n-1}, \text{ ahol } x_1, x_2, x_3 \dots x_n \text{ az idősor elemei, } n \text{ pedig az elemek száma.}$$

A tartamidősor adatainak átlagolásakor az adott időszaki egységekre jellemző adatoknak az egyszerű számtani átlagát kell számítani.



- Az átlagos változásokat kétféle mutatóval határozhatjuk meg. Az átlagos abszolút változás mutató ( $\bar{d}$ ) azt fejezi ki, hogy a vizsgált jelenség az elemzett időszak alatt időegységenként (hetente, havonta, évente stb.) átlagosan mennyivel változott. Képlete:

$$\bar{d} = \frac{x_n - x_1}{n - 1}, \text{ ahol } x_1, x_2, x_3 \dots x_n \text{ az idősor elemei, } n \text{ pedig az elemek száma.}$$

(Ez a képlet csak az időben nem ingadozó idősornál alkalmazható!)

Az átlagos relatív változás mutatója ( $\bar{l}$ ) megmutatja, hogy egy jelenség időegységenként átlagosan hány %-kal (mekkora hányaddal) változik. Képlete:

$$\bar{l} = n \sqrt[n-1]{\frac{x_n}{x_1}}, \text{ ahol } x_1, x_2, x_3 \dots x_n \text{ az idősor elemei, } n \text{ pedig az elemek száma.}$$

### ***Az idősorok analitikus vizsgálata***

Egy jelenség időbeli változásánál tapasztalati (mért) idősorral számolunk. Az általánosításhoz elméleti idősorra (valószínűségi változókra) van szükség.

Az elméleti idősor a következő összetevőkre bontható fel:

- alapirányzat vagy trend ( $\hat{y}$ )
- periodikus ingadozás ( $s$ )
- véletlen ingadozás ( $v$ )
- ciklus ( $c$ )

Az alapirányzat vagy trend az idősor alakulásának fő irányát mutatja meg, tehát egy olyan tendencia, amely maga is több tényező együttes hatását tükrözi, attól függően, hogy milyen jelenséget ír le.

Periodikus ingadozás az idősorban bizonyos okból fellépő, rendszeresen ismétlődő hullámzás. Ez lehet idényszerű vagy szezonális ingadozás.

Véletlen ingadozás az egyenként nem jelentős, egymás hatását elősegítő vagy gátló tényezők végső eredményei, amelyek a véletlen hatásának tudhatók be.

A ciklus a trend alatti vagy feletti tartósabb mozgást jelenti. Szabálytalan ingadozás, amely általában hosszabb idősoroknál figyelhető meg.

A komponensek közötti kapcsolat lehet additív, vagy multiplikatív. Ha az összetevők a lehető legegyszerűbb módon kapcsolódnak, akkor additív kapcsolatról beszélünk. A komponensek összege adja az adott időpont/időszakban ténylegesen megfigyelt értéket. Képlete:  $y = \hat{y} + s + c + v$

A multiplikatív mód esetében az összetevők szorzat adja meg az adott időpont/időszakban ténylegesen megfigyelt értéket.

$$\text{Képlete: } y = \hat{y} \cdot s \cdot c \cdot v$$

Az idősor összetevőinek additív vagy multiplikatív módját az adott idősor adatainak grafikus ábrázolásával határozhatjuk meg. A függvényen azt nézzük meg, hogy a periodikus ingadozás hullámai milyen jellegűek. Amennyiben a növekvő illetve csökkenő trend mellett nagyjából egyenlő amplitúdójúak, akkor additív a kapcsolódás, ha azonban időben növekvő vagy csökkenő tendenciát mutat a periodikus ingadozás hulláma, akkor multiplikatív a kapcsolódás.

### 13.7. A trendszámítás és módszerei

A trendszámítás célja az idősorban érvényesülő tartós tendencia meghatározása, tehát az idősor fő komponensének, a trendnek a kimutatása. Az idősorok ingadozásokat tartalmazhatnak, amelyek megnehezítik az alapirányzat, a trend felismerését. A trend meghatározásánál az alábbi módszereket szoktuk alkalmazni:

#### *Grafikus trendszámítás*

Az idősor adatait vonaldiagramban ábrázoljuk, és az alapirányzatnak megfelelő trendvonalat (pl. egyenest) közelítőleg berajzoljuk. Ez a módszer egyszerű, de pontatlan is, ezért nem gyakori ennek a használata. Célszerű megállapítani az idősor egyes tagjainál a tényszámok és a trendvonal egyes értékeinek eltérését ( $\delta = y_{ii} - y_i$ , ahol  $y_{ii}$  az  $i$  időponthoz tartozó tényszám,  $y_i$  az  $i$  időponthoz tartozó trendvonal érték. Az ugyanazon  $\delta$  érték a trendvonal különböző értékeinél más-más pontosságot jelent, ezért érdemes a  $\delta$  értékét %-ban meghatározni a következő képlettel:

$$\delta\% = \frac{\delta}{y_i} \cdot 100 = \frac{y_{ii} - y_i}{y_i} \cdot 100 \quad (\text{Minél kisebb ez az érték, annál jobban közelíti a trendvonal az értékeket.})$$

#### *Mozgó átlagolású trendszámítás*

A gyakorlatban elterjedt trendszámítási módszer, mert egyszerű és gyorsan számítható. Gyakran akkor alkalmazzuk akkor, ha nem tudjuk megadni a trendfüggvény típusát. Rövid adatsor esetén azonban ezzel a módszerrel szinte lehetetlen a trend meghatározása. Ezért először el kell döntenünk, hogy hány tagú mozgó átlagot számítunk. Ha azt tapasztaljuk, hogy szezonális van az idősor-

ban, akkor a perióduson belüli időszakok számát vagy annak többszörösét kell választani tagszámmak, hogy kisimítsa az idősor. Annyi elemből számítunk átlagot, amely a szezonális ingadozásnál átfog egy teljes idényciklust.

Az adott időszakhoz tartozó trendértéket a környező időszakok adatainak dinamikus átlagaként határozzuk meg. A mozgó átlag tagszámát ( $k$ ) úgy kell meghatározni, hogy minden esetben egy-egy idényhez tartozó adatok számával (vagy annak többszörösével) legyen egyenlő!

A mozgó átlagolású trendszámítás főbb lépései a következők:

- kiválasztjuk a mozgó átlag tagjainak számát, amit  $k$ -val jelölünk. Ezután kell kiszámítani az első  $k$  adat átlagát, ez lesz az első trend érték.
- Ezután elhagyjuk az első adatot, és kiszámítjuk az így kapott első  $k$  elem átlagát, s megkapjuk a második trendértéket.
- Így haladunk míg az utolsót is fel nem használjuk.

A kiszámított átlagértékek, amennyiben  $k$  páratlan, a kiegyenlített idősor, vagyis a trendértékek sorozata. Ha  $k$  páros szám, akkor a trendértékek meghatározásához – az elemzendő idősor konkrét időpontjaihoz/időszakaihoz hozzárendelve – egy kiegészítő lépést, (középreigazítást, vagy centírozást) kell végrehajtani. Ezt úgy oldjuk meg, hogy a kiszámított egyszerű átlagokat párosával újra átlagoljuk, és a kiszámított értékeket az átlagolt időszak közepéhez (ami most már egy konkrét időpont/időszak) rendeljük.

Páros tagszámú mozgóátlag számításának a képlete:

$$\hat{y}_t = \frac{\frac{1}{2} \cdot y_{t-k/2} + y_{t-k/2+1} + \dots + y_t + \dots + y_{t+k/2-1} + \frac{1}{2} \cdot y_{t+k/2}}{k}$$

Páratlan tagszámú mozgóátlag számításának a képlete:

$$\hat{y}_t = \frac{y_{t-(k-1)/2} + \dots + y_t + \dots + y_{t+(k-1)/2}}{k}$$

A mozgó átlagolású trendszámítás hátránya, hogy a használatánál a páratlan adatsor  $k-1$  elemmel, a páros adatsor pedig  $k$  elemmel rövidül.

### ***Analitikus trendszámítás***

Az analitikus trendszámítás a tartós tendencia meghatározásának és az előrejelzésnek az általánosan elterjedt módszere. A trendet valamely matematikai függvénnyel közelítjük, tulajdonképpen egy speciális regresszió számítást hajtunk végre.

A leggyakrabban használt függvények jellemzőit áttekintve a következőket állapíthatjuk meg:

- A lineáris trend alkalmazása az egyszerű számítása miatt nagyon elterjedt, azonban extrapolációra kevésbé alkalmas a lineáris függvény, ezért a kiterjesztésénél nagy óvatosság szükséges.
- A logaritmus függvényt regresszióra elvileg akkor használhatjuk, ha a független változó szorzatos, százalékos változására a függő változó additívan, azaz mindig ugyanannyival változik. Mivel a trendszámításnál a független változó az idő, és az sohasem változik szorzatosan, ezért ezt a függvénytypust nem használhatjuk.
- A hatványkitevős függvény jellemzője, hogy a független változó szorzatos változására a függő változó is szorzatosan reagál. Emiatt a trendeknél ez is nehezen értelmezhető, éppúgy, mint a logaritmikus függvény.
- Az exponenciális függvénynél a független változó egységnyi additív változására a függő változó szorzatosan, százalékosan reagál. Gyakorlatilag nagyon alkalmas az időben lejátszódó jelenségek leírására.
- A hiperbolák a fordított arányosság görbéi, illesztésük technikailag egyszerű, de értelmezésük már nehezebb.

A fentieknek megfelelően megkülönböztethetünk lineáris és nemlineáris trendszámítást.

Az analitikus trendszámítás alapesete a lineáris függvény, ugyanis minden egyes függvényt különböző függvény transzformációk segítségével átalakíthatunk lineáris alakra.

Mivel a lineáris trendfüggvény becsült értékeit a következő egyenlet adja:

$$\hat{y}_t = b_0 + b_1 t$$

a legkisebb négyzetek módszere így a következőket jelenti:

$$\sum (y_i - \hat{y}_t)^2 = f(b_0, b_1) = \sum (y_t - b_0 - b_1 t)^2 \Rightarrow \min$$

Az eltérésnégyzetek összegek minimumértékének meghatározása egy többváltozós szélsőérték feladat.

A szélsőérték feladat megoldásához a következő kétismeretlenes egyenletrendszer megoldása vezet, ahol az ismeretlen paraméterek a  $b_0$ ,  $b_1$ . Az egyenletrendszert normálegyenleteknek, a módszert, amellyel az egyenletekig eljutottunk legkisebb négyzetek módszerének nevezzük.

*Normálegyenletek:*

$$\begin{aligned} \sum y_t &= b_0 \cdot n + b_1 \cdot \sum t \\ \sum (t \cdot y_t) &= b_0 \cdot \sum t + b_1 \cdot \sum t^2 \end{aligned}$$

A legkisebb négyzetek módszer alkalmazásával a fenti egyenletrendszer  $b_0$  és  $b_1$  gyökei adják a becsült lineáris trend együtthatóinak értékét.

*Paraméterek értelmezése:*

$b_1$ : (függvény meredeksége) a trendfüggvény várható értékére megadja az időszakonkénti állandó abszolút változás nagyságát, azaz azt jelenti, hogy a  $t$  időegységnyi változására mennyivel változik a trend értéke.

$b_0$ : a  $t=0$  időszak/időponthoz tartozó trendértéket adja meg.

Amennyiben az idősorhoz tartozó időpontokat/időszakokat nem a vizsgált időszak elejétől sorszámozzuk, hanem az idősor közepétől kezdjük az elemzést, (tehát az időszak közepe előtti időpontok/időszakokhoz negatív sorszám kerül), akkor az ún.  $\Sigma t=0$  módszer szerinti elemzést hajtjuk végre.

Ha páratlan elemű az idősorunk, akkor egyszerűen az idősor középső eleméhez a nulla sorszámot rendeljük, és időben visszafelé haladva egy egységgel csökken a sorszám, időben előre tovább haladva pedig nő egy egységgel a sorszáma.

A trend paramétereinek becslése  $\Sigma t=0$  módszer esetén:

(Az egyszerűsítésekből adódóan a normálegyenletek leegyszerűsödnek)

$$\sum y_t = b_0 \cdot n \Rightarrow b_0 = \frac{\sum y_t}{n}$$

$$\sum (t \cdot y_t) = b_1 \cdot \sum t^2 \Rightarrow b_1 = \frac{\sum (t \cdot y_t)}{\sum t^2}$$

A paraméterek alapértelmezése nem változik.

$b_1$ : megmutatja, hogy a  $t$  egy időegységnyi változására a trend mennyivel változik. (Arra kell odafigyelni, hogyha az idősor páros elemszámú volt, és  $\Sigma t=0$  módszert alkalmaztuk, akkor a  $t$  egy időegysége nem azonos a "valós" időegységgel.)

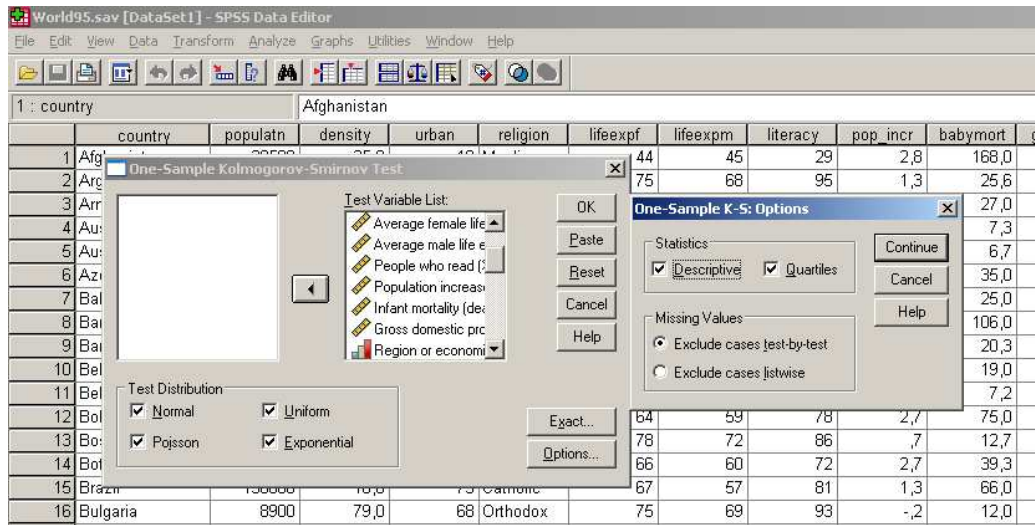
$b_0$ : a  $t=0$  időszakhoz rendelt trendérték. (Ha a  $\Sigma t=0$  módszert alkalmaztuk, akkor ez az érték az idősor közepéhez tartozó trendérték, ha páratlan elemszámú az idősor, egyébként nem rendelhető hozzá konkrét időpont/időszak. Ez a trendérték nem más, mint az idősorban lévő összes érték számtani átlaga, mivel

$$\sum y_t = \sum \hat{y}_t$$

### 13.8. SPSS alkalmazások

Készítsünk korrelációs mátrixot a **World95.sav** állomány változóit felhasználva. Ehhez először az adatsorok eloszlástípusának vizsgálatát kell megejteni. Az **Analyze - Nonparametric test - One Sample K-S** próbát kiválasztva normál, egyenletes, poisson és exponenciális eloszlásra tesztelhetjük az adatsort (binomiálisra nem, ezért kvázinormál-eloszlásúnak fogja őket venni). Az opciók-

nál kiválaszhatjuk, hogy kérünk-e leíró statisztikát. Ekkor az alapvető statisztikai paramétereket, pl. a mediánt és a kvartiliseket is megkapjuk.



		Population in thousands	Number of people / sq. kilometer	People living in cities (%)	Average female life expectancy	Average male life expectancy	People who read (%)	Population increase (% per year)	Infant mortality (deaths per 1000 live births)	Gross domestic product / capita
N		109	109	108	109	109	107	109	109	109
Normal Parameters <sup>a,b</sup>	Mean	47723,88	203,415	56,53	70,16	64,92	78,34	1,682	42,313	5859,98
	Std. Deviation	146726,364	675,7052	24,203	10,572	9,273	22,883	1,1976	38,0792	6479,836
Most Extreme Differences	Absolute	,373	,383	,090	,174	,164	,172	,097	,169	,204
	Positive	,323	,356	,063	,131	,116	,172	,097	,169	,204
	Negative	-,373	-,383	-,090	-,174	-,164	-,171	-,058	-,157	-,188
Kolmogorov-Smirnov Z		3,896	3,999	,932	1,817	1,713	1,778	1,010	1,769	2,135
Asymp. Sig. (2-tailed)		,000	,000	,350	,003	,006	,004	,259	,004	,000

a. Test distribution is Normal.

b. Calculated from data.

		Population in thousands	Number of people / sq. kilometer	People living in cities (%)	Average female life expectancy	Average male life expectancy	People who read (%)	Population increase (% per year)	Infant mortality (deaths per 1000 live births)	Gross domestic product / capita
N		109	109	108	109	109	107	109	109	109
Uniform Parameters <sup>a,b</sup>	Minimum	256	2,3	5	43	41	18	-,3	4,0	122
	Maximum	1205200	5494,0	100	82	76	100	5,2	168,0	23474
Most Extreme Differences	Absolute	,839	,870	,137	,368	,346	,374	,286	,420	,438
	Positive	,839	,870	,056	,009	,028	,009	,286	,420	,438
	Negative	-,009	-,009	-,137	-,368	-,346	-,374	-,026	-,009	-,009
Kolmogorov-Smirnov Z		8,758	9,078	1,423	3,839	3,615	3,866	2,986	4,384	4,570
Asymp. Sig. (2-tailed)		,000	,000	,035	,000	,000	,000	,000	,000	,000

a. Test distribution is Uniform.

### Eloszlásvizsgálat Kolmogorov-Szmirnov próbával az eredeti adatsoron normál és egyenletes eloszlásokra

A kapott táblázatból a szignifikancia-értéket (**Asymp. sig.**) leolvastva eldönthetjük, hogy adatsoraink eloszlása mely eloszlásra hasonlít leginkább. Példánkban a városban lakók aránya, s a népességnövekedés mutat csak 0,05-nél nagyobb értéket, azaz közelíthető normáeloszlással. Még így is előfordulhat,

hogy egy másik eloszlás szignifikanciaszintje magasabb lesz. Jelen állapotában az adatsorral csak Spearman-féle rangkorrelációt (**nonparametric**) számíthatunk, Pearson-félét nem, mert az utóbbihoz normáeloszlású, arány-vagy intervallumskálán mért adatokra van szükség, melyek között lineáris kapcsolat van. Az **Analyze - Correlate - Bivariate** parancsot kiválasztva lehetőségünk van a korrelációs mátrix előállítására tetszőleges számú változó felhasználásával.

Nem használhatunk olyan adatokat, melyek számként definiáltak ugyan, de valójában csak csoportok elkülönítésére használtuk őket (pl. Debrecen=1, Szeged=2), ha a csoportok között nincs minőségi különbség.

A Spearman-féle korreláció ordinális adatokat, rangsorrendet használ.

Minél több adatpár szerepel a korrelálandó adatsorokban, annál megbízhatóbb a korrelációs vizsgálat (A szignifikáns korrelációs koefficienseket jelöli a program: **Flag significant...**).

			Correlations					
			Population in thousands	Number of people / sq. kilometer	People living in cities (%)	Population increase (% per year)	Gross domestic product / capita	Aids cases
Spearman's rho	Population in thousands	Correlation Coefficient	1,000	,171	-,103	-,077	-,183	-,278**
		Sig. (2-tailed)	.	,076	,288	,425	,057	,004
		N	109	109	108	109	109	106
	Number of people / sq. kilometer	Correlation Coefficient	,171	1,000	-,022	-,312**	,037	-,050
		Sig. (2-tailed)	,076	.	,824	,001	,699	,611
		N	109	109	108	109	109	106
	People living in cities (%)	Correlation Coefficient	-,103	-,022	1,000	-,427**	,735**	-,025
		Sig. (2-tailed)	,288	,824	.	,000	,000	,797
		N	108	108	108	108	108	105
	Population increase (% per year)	Correlation Coefficient	-,077	-,312**	-,427**	1,000	-,608**	-,111
		Sig. (2-tailed)	,425	,001	,000	.	,000	,256
		N	109	109	108	109	109	106
	Gross domestic product / capita	Correlation Coefficient	-,183	,037	,735**	-,608**	1,000	,037
		Sig. (2-tailed)	,057	,699	,000	,000	.	,704
		N	109	109	108	109	109	106
	Aids cases	Correlation Coefficient	-,278**	-,050	-,025	-,111	,037	1,000
		Sig. (2-tailed)	,004	,611	,797	,256	,704	.
		N	106	106	105	106	106	106

\*\* . Correlation is significant at the 0.01 level (2-tailed).

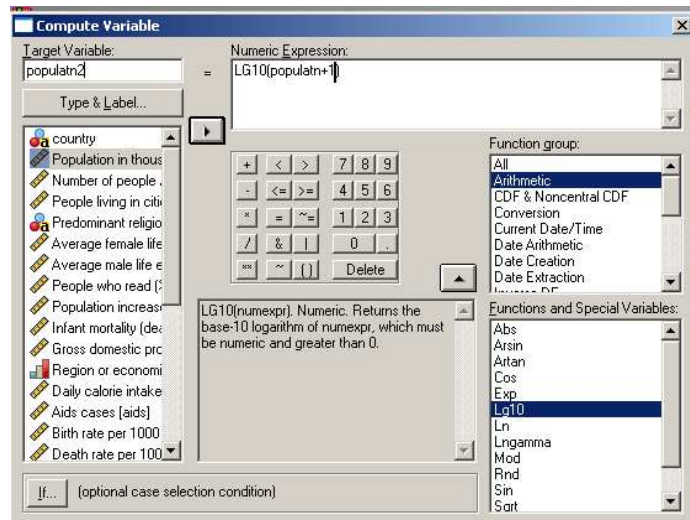
### Spearman-féle rangkorreláció

Pearson-féle korreláció számításához az adatsorok eloszlásán módosítani kell, hogy normáeloszlást mutassanak. Vagy intervallumokba soroljuk az adatokat, de ekkor gondoskodni kell arról, hogy az intervallumok száma kellően nagy legyen, hogy a kvázi-normáeloszlás értelmezhető legyen, vagy pedig az  $\lg(x+1)$  képletet használjuk. Az első esetben a **Transform - Recode into new variables** parancsot használhatjuk (utána érdemes egy hisztogramot kérni, hogy az újradefiniált adatsor eloszlása hasonlít-e a normáeloszláshoz: **Graph - Legacy Dialogs - Histogram**), az utóbbi esetben a **Transform - Compute Variable** lesz a megoldás.

Példánkban az utóbbi megoldást választjuk és vetjük össze az eredeti adatsor Pearson-féle korrelációs mátrixát a logaritmizált értékek korrelációs mátrixával. A logaritmust az aritmetikus függvényeken belül leljük meg, az átalakított értékeket új változónévvel mentjük el. Ezt követően ismét tesztelni kell az

eloszlást Kolmogorov-Szmirnov próbával, ugyanis az  $\lg(x+1)$  nem minden esetben képes normáloszlásúvá tenni az adatsort (pl. a várható női és férfi átlagéletkor esetében nem, a népességszám, GDP/fő, népsűrűség esetében igen!).

Ezt követően készítjük el a Pearson-féle korrelációs mátrixot.



One-Sample Kolmogorov-Smirnov Test

		populatr2	number_of_density2	Log (base 10) of GDP_CAP	Log (base 10) of AIDS_RT	People living in cities (%)	Population increase (% per year)
N		109	109	109	106	108	109
Normal Parameters <sup>a,b</sup>	Mean	4,1141	1,8021	3,4218	1,3800	56,53	1,682
	Std. Deviation	,65402	,59986	,62006	,70943	24,203	1,1976
Most Extreme Differences	Absolute	,093	,060	,085	,085	,090	,097
	Positive	,093	,060	,074	,085	,063	,097
	Negative	-,060	-,055	-,085	-,053	-,090	-,058
Kolmogorov-Smirnov Z		,969	,624	,882	,876	,932	1,010
Asymp. Sig. (2-tailed)		,305	,830	,417	,427	,350	,259

a. Test distribution is Normal.

b. Calculated from data.

Eloszlásvizsgálat Kolmogorov-Szmirnov próbával a logaritmizált adatsoron normáloszlásra

Correlations

		Population in thousands	Number of people / sq. kilometer	People living in cities (%)	Population increase (% per year)	Gross domestic product / capita	Aids cases
Population in thousands	Pearson Correlation	1	-,018	-,175	-,050	-,087	,138
	Sig. (2-tailed)		,850	,071	,603	,370	,157
	N	109	109	108	109	109	106
Number of people / sq. kilometer	Pearson Correlation	-,018	1	,223*	-,165	,201*	-,042
	Sig. (2-tailed)	,850		,020	,086	,036	,668
	N	109	109	109	109	109	106
People living in cities (%)	Pearson Correlation	-,175	,223*	1	-,375**	,605**	,043
	Sig. (2-tailed)	,071	,020		,000	,000	,660
	N	108	108	108	108	108	105
Population increase (% per year)	Pearson Correlation	-,050	-,165	-,375**	1	-,521**	-,056
	Sig. (2-tailed)	,603	,086	,000		,000	,571
	N	109	109	108	109	109	106
Gross domestic product / capita	Pearson Correlation	-,087	,201*	,605**	-,521**	1	,258**
	Sig. (2-tailed)	,370	,036	,000	,000		,008
	N	109	109	108	109	109	106
Aids cases	Pearson Correlation	,138	-,042	,043	-,056	,258**	1
	Sig. (2-tailed)	,157	,668	,860	,571	,008	
	N	106	106	105	106	106	106

\*. Correlation is significant at the 0.05 level (2-tailed).

Pearson-féle korrelációs mátrix az eredeti adatsort használva



Correlations

		populatn2	number_of_density2	People living in cities (%)	Population increase (% per year)	Log (base 10) of GDP_CAP	Log (base 10) of AIDS_RT
populatn2	Pearson Correlation	1	,136	-,138	-,078	-,217*	-,206*
	Sig. (2-tailed)		,157	,155	,418	,023	,034
	N	109	109	108	109	109	106
number_of_density2	Pearson Correlation	,136	1	,022	-,256**	,013	-,134
	Sig. (2-tailed)	,157		,822	,007	,896	,171
	N	109	109	108	109	109	106
People living in cities (%)	Pearson Correlation	-,138	,022	1	-,375**	,754**	-,061
	Sig. (2-tailed)	,155	,822		,000	,000	,539
	N	108	108	108	108	108	105
Population increase (% per year)	Pearson Correlation	-,078	-,256**	-,375**	1	-,557**	-,066
	Sig. (2-tailed)	,418	,007	,000		,000	,502
	N	109	109	108	109	109	106
Log (base 10) of GDP_CAP	Pearson Correlation	-,217*	,013	,754**	-,557**	1	,057
	Sig. (2-tailed)	,023	,896	,000	,000		,563
	N	109	109	108	109	109	106
Log (base 10) of AIDS_RT	Pearson Correlation	-,206*	-,134	-,061	-,066	,057	1
	Sig. (2-tailed)	,034	,171	,539	,502	,563	
	N	106	106	105	106	106	106

\*. Correlation is significant at the 0.05 level (2-tailed).

\*\*. Correlation is significant at the 0.01 level (2-tailed).

Pearson-féle korrelációs mátrix a logaritmizált adatsort használva

Látható, hogy van különbség a két adatsor korrelációs értékei között, miképp a Spearman-féle rangkorreláció korrelációs koefficiensei is eltérnek: például a városban lakók aránya és az egy főre jutó GDP között a korrelációs koefficiens értéke 0,735 Spearman-féle korrelációt használva, 0,605, ha nem normáloszlású adatsorokat korreláltunk és 0,754, ha normáloszláshoz hasonló eloszlást mutató adatsorokat korreláltunk Pearson-féle korrelációt használva.

### Parciális korreláció

Kettőnél több változó esetén felvetődhet annak a lehetősége, hogy a és b változók között a kapcsolat azért olyan erős, mert a és b c-vel is erős kapcsolatot mutat (ez a *multikollinearitás* jelensége), avagy ettől függetlenül van meg az összefüggés a két változó között. Tehát a és b kapcsolata lehet közvetett és közvetlen is. E kérdés eldöntéséhez *parciális korrelációs* vizsgálatot szoktunk végezni.

A parciális korrelációval azt vizsgáljuk, hogy két változó kapcsolatát hogyan befolyásolja a harmadik, vagy több változó korrelációs koefficiensének értéke. A parciális korreláció lényegének érzékeltetésére a következő példát lehet felhozni.

Van három adatsorunk: az egyik a kocsmai verekedések számát mutatja, a másik az összetört poharak számát, a harmadik a rendőri megjelenések számát. A rendőrök száma és az összetört poharak száma között 0,8-as korrelációs koefficiens mérünk. Vajon ez azt jelenti, hogy a rendőrök törik össze a poharakat? Az előbbieket értelmében egy korrelációs koefficiens alapján nem következtethetünk ok-okozati viszonyra. Ha megvizsgáljuk a rendőrök megjelenését és a pohártörések számát, kapcsolatát a kocsmai verekedések számának függvényében (ez lesz

az ún. kontroll, **Analyze – Correlate – Partial, control variable**) és azt találjuk, hogy a korrelációs koefficiens leesett pl. 0,5-re, akkor egyértelmű, hogy a kapcsolat erősségének alakításában a verekedések száma is részt vesz. Másképpen fogalmazva: nem a rendőrök törik a poharakat. (Sokszor nincs lehetőségünk ilyen explicit kijelentéseket tenni, pl. természetföldrajzi változók esetében). Ha azt találjuk, hogy a kapcsolat erőssége 0,8 körül marad, akkor a verekedések száma nincs hatással a másik két változó kapcsolatára, tehát a rendőrök törik a poharakat. A vizsgálathoz értelemszerűen az eredeti korrelációs mátrixra is szükség van, hiszen a parciális korrelációs értékeket ezekkel vetjük össze.

A **parciális korreláció** számításához válasszunk ki két változót, melynek ismert a korrelációs koefficiense. Legyen ez a városokban lakók aránya és az egy főre jutó GDP (annak logaritmizált adatsora). Ekkor a korrelációs koefficiens értéke 0,754, tehát erős kapcsolatra utal a két változó között. Nézzük meg, hogy a többi változó hogyan befolyásolja ezt a kapcsolatot. Az **Analyze - Correlate - Partial** parancsot választva kiválasztható a két változó, amelynek kapcsolatát vizsgáljuk, kontrollként (**Controlling for**) pedig egy vagy több változó. A népeség szám például nem befolyásolja a két változó kapcsolatának erősségét, mert a korrelációs koefficiens 0,749-es értéket mutat. Ugyanígy nem befolyásolja ezt a kapcsolatot az AIDS-esek száma, a népsűrűség, az éves szaporulat és a népesség együttesen.

1	Afghanistan	20500	25,0	18	Muslim	44	45
2	Argentina	33900	12,0	86	Catholic	75	68
3	Armenia	3700	126,0	68	Orthodox	75	68
4	Australia	19000	20,0	80	Other	80	74
5	Austria	8000	20,0	79	Other	79	73
6	Azerbaijan	7000	20,0	75	Other	75	67
7	Bahrain	1000	20,0	74	Other	74	71
8	Bangladesh	10000	20,0	53	Other	53	53
9	Barbados	2000	20,0	78	Other	78	73
10	Belarus	10000	20,0	76	Other	76	66
11	Belgium	10000	20,0	79	Other	79	73
12	Bolivia	10000	20,0	64	Other	64	59
13	Bosnia and Herzegovina	10000	20,0	78	Other	78	72
14	Botswana	10000	20,0	66	Other	66	60
15	Brazil	10000	20,0	67	Other	67	57
16	Bulgaria	10000	20,0	75	Other	75	69
17	Burkina Faso	10000	20,0	50	Other	50	47
18	Burundi	10000	20,0	50	Other	50	46
19	Cambodia	10000	20,0	52	Buddhist	52	50
20	Cameroon	13100	27,0	40	Animist	58	55

### Correlations

Control Variables			People living in cities (%)	Log (base 10) of GDP_CAP
populatr2 & number_of_density2 & Population increase (% per year) & Log (base 10) of AIDS_FIT	People living in cities (%)	Correlation	1,000	,694
		Significance (2-tailed)	.	,000
		df	0	99
	Log (base 10) of GDP_CAP	Correlation	,694	1,000
		Significance (2-tailed)	,000	.
		df	99	0

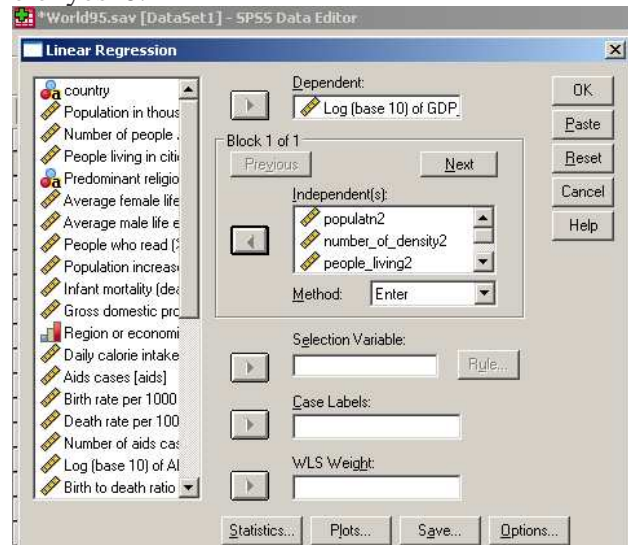
### ***Többváltozós (multivariate) lineáris regresszió az SPSS-ben***

A parciális korrelációs vizsgálatokból továbblépést jelent az úgynevezett regresszió-**analízis**. A parciális korreláció és a regresszió-analízis között van különbség. Az előbbi esetben egy változópár kapcsolatának befolyásolásáról van szó több másik tényező hatására, míg a regresszió-analízis lényege, hogy megmutatja, hogy a változók értékeinek (**independent variable**) alakulása hogyan hat egy kiválasztott változó (**dependent variable**) értékeire, azaz e változók mennyire magyarázzák, tükrözik vissza annak értékeit, melyik az a változó, amelyik leginkább befolyásolja/meghatározza egy másik változó értékét.

Az előbbi példák mintájára meghatározható, hogy pl. a GDP/fő vagy a népességszám alakításában mely tényezők vesznek részt. Ekkor regresszió-analízist végzünk. (Itt sem hátrány, ha az adatsorok normáleloszlást mutatnak) Kiválasztjuk az **Analyze - Regression - Linear** parancsot, majd a vizsgálni kívánt változót, végül azokat a változókat, melyek hatását, hatásának erősségét vizsgálni kívánjuk. Választhatjuk az **Enter, Remove, Forward, Backward** stb. módszereket. Az enter egyszerre vizsgálja a tényezőket, a forward egyenként tárja fel szerepüket a kapcsolat alakításában. A Statistics menüben végezhetünk alapstatisztikai próbákat (Durbin-Watson, kollinearitás-diagnózis). Az adatok elmenthetők (**Save**).

Ha a regressziós modellben szereplő változók egymással korrelálnak, összefüggnek, akkor a független, magyarázó változók (independent variable) nemcsak a függő változót, de egymást is magyarázzák (**multikollinearitás**), ami sok értelmetlen eredményt adhat. Célszerű tehát egymással nem korreláló változókat használni, ami nem könnyű, de a főkomponens-analízis, faktoranalízis éppen erre való.

Az spo file-ban az a **Modell Summary** tábla tartalmazza az  $R^2$  értékét, a **Coefficients** tábla a *standardizált beta* értékeket. Az  $R^2$  azt jelenti, hogy a modellben szereplő magyarázó változók (independent variables), a függő változó értékeinek alakulását hány %-át magyarázza meg (pl.  $R^2=0,63$  63%-ban). A standardizált beta azt mutatja, hogy az egyes magyarázó változók közül melyik szerepe a legkiemelkedőbb. Esetünkben a Standardizált beta koefficiens a városokban lakók számánál a legnagyobb, tehát az utóbbi befolyásolja leginkább adathalmazunk GDP értékeit (ami nem meglepő, hiszen a korrelációs vizsgálat nagy korrelációs koefficienst mutatott ki, a parciális korreláció pedig bizonyította, hogy e két változó közötti kapcsolatot a többi változó alig befolyásolja). A többi tényező szerepe elenyésző.



ANOVA<sup>a</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	27,050	5	5,410	38,047	,000 <sup>a</sup>
	Residual	14,077	99	,142		
	Total	41,128	104			

a. Predictors: (Constant), Population increase (% per year), Aids cases, populatin2, number\_of\_density2, people\_living2

b. Dependent Variable: Log (base 10) of GDP\_CAP

Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients		Sig.
		B	Std. Error	Beta	t	
1	(Constant)	2,328	,435		5,349	,000
	populatin2	-,205	,059	-,217	-3,488	,001
	number_of_density2	-,031	,065	-,030	-,482	,631
	people_living2	1,346	,159	,551	8,461	,000
	Aids cases	2,08E-006	,000	,136	2,214	,029
	Population increase (% per year)	-,188	,035	-,360	-5,383	,000

a. Dependent Variable: Log (base 10) of GDP\_CAP

## Feladatok

- 1, Túránk során 8 pontot érintettünk, s mindegyik távolsága ismeretes egymástól (km). Határozzuk meg a pontok elhelyezkedését egymáshoz képest (SPSS-ben).

	1	2	3	4	5	6	7	8
1	0							
2	8	0						
3	3	9	0					
4	4	8	11	0				
5	6	6	6	14	0			
6	12	13	7	2	2	0		
7	3	5	4	7	3	8	0	
8	1	2	3	8	5	9	2	0

- 2, Szociálgeográfiai vizsgálatunkban 8 néprajzi régió egymáshoz való viszonyát minősítettük egy skálán 1-10 között (a legjobb kép az 1-es, legrosszabb viszony a 10-es értéket kapta) Határozzuk meg a néprajzi régiók közötti belső csoportok elhelyezkedését egymáshoz képest!

	matyók	barkók	palócok	kunok	jászok	hajdúk	bunyevácok	tírpákok
matyók	1							
barkók	8	1						
palócok	8	7	1					
kunok	4	5	3	1				
jászok	5	4	4	2	1			
hajdúk	5	3	4	4	3	1		
bunyevácok	3	2	7	6	4	2	1	
tírpákok	2	5	8	6	6	9	2	1

- 3, Egy elektronikus távmérővel különböző nagyságú, ismert valódi értékű  $d_i$  távolságokat mértünk. ( $d = \dots$  a mért távolságok 0,1 km élességgel és a mért távolságok  $\Delta_i$  valódi hibái:  $\dots$   $|\Delta_i|$  abszolút értékei:  $\dots$ ). Vizsgáljuk meg, találunk-e *lineáris összefüggést* a  $d_i$  távolságok nagysága és a  $|\Delta_i|$  értékek között. Határozzuk meg a korrelációs együtthatót és írjuk fel a regressziós egyenes egyenletét.
- 4, Egy városban 10600 nő él, ők teszik ki a lakosság 53 %-át. A nőkön belül a nyugdíjasok aránya 35 %, a férfiaknál ugyanez az arány 25 %.
- Számítsa ki a függetlenség esetén várható gyakoriságokat!
  - Állapítsa meg, milyen szoros az ismérvek között az összefüggés!
- 5, Egy régió 400 háztartásának megoszlása a hűtőszekrény típusa és a település jellege szerint:

A település jellege	A hűtőszekrény típusa		Összesen
	Hagyományos	Fagyasztó	
Város	160	30	190
Község	120	90	210
Összesen	280	120	400

Állapítsa meg, van-e kapcsolat a két adatsor között!

**6, Munkanélküliek száma nemek és foglalkozási főcsoportok szerint egy régióban, 2004**

*ezer fő*

Foglalkozási főcsoportok	Férfi	Nő	Összesen
Felsőfokú végzettségű	10,6	11,3	21,9
Ügyviteli és szolgáltatásban	14,9	24,5	39,4
Szakképzett fizikai	53,1	20,1	73,2
Szakképzetlenek	24,8	11,9	36,7
Összesen	103,4	67,8	171,2

Határozza meg a kapcsolat szorosságának mérőszámát!

**7, A Magyarországra érkező külföldiek száma havonként, ezer fő:**

Hónapok	2000	2001	2002	2003
Január	1 562	1 824	1 594	1 730
Február	1 548	1 792	1 626	1 669
Március	2 000	2 055	2 209	2 132
Április	2 550	2 689	2 345	2 509
Május	2 636	2 658	2 702	2 631
Június	2 739	2 804	2 754	2 850
Július	3 666	3 568	3 803	3 663
Augusztus	4 229	4 130	4 248	4 226
Szeptember	2 824	2 760	2 957	2 840
Október	2 640	2 475	2 748	2 708
November	2 323	2 100	2 379	2 202
December	2 424	1 823	2 375	2 252
Év, összesen	31 141	30 679	31 739	31 412

Végezze el az idősorok összetevőinek elemzését!

**8, A népességszám és a születések számának alakulása hazánkban:**

Év	Népesség az év elején, ezer fő	Élveszületések száma
1997	10301	100350
1998	10280	97301
1999	10253	94645
2000	10222	97597
2001	10200	97047
2002	10175	96804
2003	10142	94647

Elemesse a népességszám- és az élveszületések száma változásának alapirányzatát mozgóátlagolással és analitikus trendfüggvénnyel!

**9, Néhány véletlen mintaként kiválasztott család vonatkozásában a rendelkezésre álló jövedelem és fogyasztás egy főre vetített havi adatai és azok ábrája az alábbiak:**

Jövedelem (ezer Ft)	Fogyasztás értéke, ezer Ft
39,8	38,6
43,1	44,4
46,2	40,4
49,7	45,9
52,8	50,5
54,8	44,8
56,1	49,5
61,6	53,2
59,7	51,4

a, Határozza meg a jövedelem függvényében változó fogyasztás regressziós modelljét, a kapcsolat-szorossági mutatót az alábbi részszámítások ismeretében és vonjon le következtetéseket!

b) Becsülje meg a 60 ezer Ft-os jövedelem szinthez tartozó fogyasztási értéket!

**10, A csapadékos napok és az extrém talajvízállású napok közötti kapcsolatot vizsgálva a következő számpárokat kaptuk (3;5, 5;3, 10;10, 15;15, 17; 14, 20;8, 25;20, 21;23, 21;16, 30;30). Ábrázolja őket kétváltozós koordináta-rendszerben, végezze el a normalitás-vizsgálatot (Kolmogorov-Szmirnov**

próba), határozza meg a korrelációt (amennyiben az adatok normáeloszlásúak, határozza meg a Pearson korreláció mellett a rangkorreláció koefficiensét is), illesszen a pontokra regressziós görbét a legnagyobb pontossággal és határozza meg az egyenletét.

**11,** „A” kőzet keménysége 5 minta átlaga alapján 20 MPa, a rajta kiformalódott lejtő átlagos lejtése 14%, „B” kőzeté 23 MPa, lejtői átlagosan 16%-os lejtést mutatnak, „C” kőzet 8 MPa és 10% értékekkel jellemezhető, „D” kőzet 55 MPa és 19%, „E” kőzet: 66 MPa és 23%, „F” kőzet esetében az értékpár: 79 MPa és 26%, G: 78 MPa, 25%, H: 123 MPa, 23%, I: 145 MPa, 26%, J: 250 MPa 29%, K: 4 MPa, 5%. Ábrázolja őket kétváltozós koordináta-rendszerben, végezze el a normalitás-vizsgálatot (Kolmogorov-Szmirnov próba), határozza meg a korrelációt, illesszen a pontokra regressziós görbét a legnagyobb pontossággal és határozd meg az egyenletét.

**12,** Vizsgálja meg a távolság függvényében a különböző szemcsekategóriák (összegezze a homok, iszap, agyagként az egyes részfrakciókat) előfordulási gyakorisága között a kapcsolat erősségét!

*(Figyelem! A vizsgálatot az azonos településről és nagyjából hasonló tszf. magasságból vett minták esetében végezze el először, majd vonja be a vizsgálatba a többi települést is. Mennyire változik a korreláció? Településtől független-e a szemcsefrakciók gyakorisága?)*

Szemcse Ø mm-ben Mintavétel helye	Homok					Kőzetliszt (Iszap)			Agyag	%
	Közép- szemű >0,32	0,32-0,2	Apró- szemű 0,2-0,1	Finom- Szemű 0,1-0,05	Por 0,05-0,02	0,02-0,01	0,01-0,005	0,005-0,002	<0,002	
1. Gulács, Szilvás-szeg jobb parti hát 0-25 cm, 5m-re a medertől	0,1	0,3	4,8	3,5	25,6	30,6	17,4	10,1	7,6	100
2. Gulács, Szilvás-szeg jobb parti hát 25-40 cm, 5m-re a medertől	<b>6,3</b>	<b>26,3</b>	<b>41,8</b>	4,7	6,5	4,7	3,9	<b>3,0</b>	<b>2,8</b>	100
3. Gulács, Szilvás-szeg jobb parti 0-30 cm, 30 m-re a medertől	0,1	0,2	6,1	14,2	30,2	22,2	10,1	9,1	7,8	100
4. Tarpa, a gát hullámtéri oldala 200 m-re a medertől	0,2	0,2	10,6	11,6	36,5	17,8	9,2	7,5	6,4	100
5. Gulács, Boroszló-kert hullámtér 1 km-re a medertől	-	-	-	4,2	16,6	22,2	24,6	16,0	16,4	100
6. Gulács, védett ártér, mélyebb felszín a gátszakadástól 6 km-re	2,2	0,5	0,5	10,5	33,9	20,1	14,6	8,8	8,9	100
7. Tákos, 1. sarlólapos a gátszakadástól 12 km-re	-	-	0,6	3,2	10,9	7,1	17,7	<b>31,3</b>	<b>29,2</b>	100



hónap	vízszint (m)	csapadék (mm)	hőmérséklet
1	-2,14	24	-1
2	-2,22	22	0
3	-0,45	98	2,3
4	-0,69	107	5,6
5	-1,79	40	9,6
6	-1,66	45	16,7
7	-1,98	34	18,4
8	-3,14	2	27,8
9	-2,99	23	24,5
10	-1,56	85	15,2
11	-1,78	62	5,4
12	-2,45	30	1,6

- 13,** A fenti táblázat alapján határozza meg a korrelációs koefficiens értékét a változók között!
- 14,** A fenti táblázat adatain végezzen parciális korrelációs vizsgálatot minden változóra!
- 15,** Határozza meg és ábrázolja diagramon, hogy a 3 talajvízkútban a vízszint mozgása mennyire követi egymást, majd számolja ki a korrelációs koefficiens a 3 kút vízszintingadozása között!

hónap	1. talajvízkút		2. talajvízkút		3. talajvízkút	
	vízszint (m)	csapadék (mm)	vízszint (m)	csapadék (mm)	vízszint (m)	csapadék (mm)
1	-2,34	24	-2,14	29	-2,14	22
2	-2,45	22	-2,45	23	-2,22	98
3	-1,77	92	-2,22	33	-0,45	107
4	-1,11	87	-1,4	42	-0,69	40
5	-1,45	66	-1,02	88	-1,79	45
6	-1,79	69	-0,67	113	-1,66	34
7	-2,56	32	0	78	-1,98	2
8	-2,59	45	-1,1	45	-3,14	23
9	-2,4	60	-1,6	23	-2,99	85
10	-2,7	41	-1,3	46	-1,56	62
11	-2,56	45	-1,6	44	-1,78	24
12	-2,22	48	-2,1	23	-2,45	30

- 16,** A talajvízszint ingadozása követi-e a lehulló csapadékmennyiséget? A korrelációs koefficiens mutat-e változást, ha a késleltető hatással számolva eltoljuk a talajvíz és csapadék adatokat egymáshoz képest?



## 14. Faktoranalízis, főkomponens-analízis

A földtudományok adatbázisaiban gyakran több tíz adatfészeség áll rendelkezésünkre objektumonként. Az ilyen sokdimenziós az adatrendszer megnehezíti a klaszterezést (csoportképzést, lásd következő fejezet), és elfogadhatatlan mértékben megnövelheti a számítási időt. Ezért a matematikai statisztikában kidolgoztak olyan eljárásokat, amelyek csökkentik az adatrendszer dimenziószámát, ugyanakkor minimalizálják az így elvesztett információt. Ezekkel a módszerekkel ismerkedünk a következő fejezetekben.

### 14.1. Faktoranalízis

A *faktoranalízis* egy olyan adatelemzési eljárás, amelynek a lényege a változók csoportosítására, és az adatok *redukciója (dimenziótlantítás)*. Ha túl sok változót vontunk be a vizsgálatba, akkor számítógépes adatfeldolgozásnál szükséges lehet a fájl méretének csökkentése, a változók számának redukciója, vagy pedig a változók függetlenségének vizsgálata.

Célja lehet tehát:

- a, a változók számának redukciója
- b, a klaszteranalízis, többváltozós regresszió megkönnyítése egymással nem korreláló változócsoporthoz létrehozásával
- c, független változók kiválogatása egy változóhalmazból, melyek nagy korrelációt mutatnak a faktorokkal

A faktoranalízis egy olyan elemző módszernek tekinthető, amellyel egy többváltozós összefüggésrendszer háttérváltozóit kívánjuk feltárni. A faktoranalízis alkalmazásával az a célunk, hogy az adatbázisunk változóit olyan faktorokkal fejezzük ki, mely faktorok az összefüggő változók közötti kapcsolatot jellemzik, tehát az eredeti változók szórásának túlnyomó része megmagyarázható.

Két vagy több tetszőleges valószínűségi változó közötti korreláció létrejötte elképzelhető közös keletkezési feltételek alapján. Ezeket a közös keletkezési feltételeket nevezzük faktoroknak, melyek egymástól függetlenek, azaz egymással nem korrelálnak. A korrelációs együtthatók mátrixot képeznek, és ezekből a korrelációs együtthatókból a faktorok megbecsülhetők.

A faktoranalízis alkalmazásával elérhetjük azt is, hogy a változók felhasználásával olyan összefüggéseket tudunk azonosítani, amelyek közvetlenül nem figyelhetők meg. Sok jelenség, folyamat önmagában nem figyelhető/mérhető, de több különböző adat megjelenéséből lehet a létezésére következtetni.

A faktorelemzést használhatjuk olyan tényezők vagy faktorok azonosítására, amelyek az adott változók közötti korrelációt magyarázzák. Előfordul az is, hogy kevesebb számú korrelálatlan változót azonosítunk, amelyekkel az eredeti

korrelált változók helyettesíthetőek a későbbi többváltozós elemzések során. Alkalmazható néhány kiemelkedően fontos változó azonosítására is, amelyek később a többváltozós elemzésekhez felhasználhatók.

A faktoranalízis során valamennyi változónak kiszámítjuk a korrelációját. Ebből látható, hogy mely változókat nem vesszük figyelembe, mert a többi változóval nagyon kicsi a korrelációjuk. Közös faktorok csak olyan változók számára léteznek, amelyek relatív jól korrelálnak egymással.

Az egyes faktorok kiválasztását a sajátérték meghatározása segíti, amely mutatja, hogy az adott faktor(ok) az eredeti változók teljes varianciáját mennyiben magyarázzák meg. A kiszűrendő faktorok számának behatárolására használjuk. Maximálisan annyi sajátérték számítható ki, amennyi a változók száma. A sajátértékek nagyság szerinti rendezése után azonban az utolsó faktorok olyan kis súlyúak, hogy már nem tartalmaznak lényeges információkat. A kiszűrendő faktorok számához ezért korlátot kell megadni. Ezen korlát megadásához azonban nincsenek szigorú feltételek, vagy teljesen egyértelmű megoldások. Tapasztalati értékek alapján legáltalánosabb a  $1=1$  korlát használata, azaz az 1-nél nagyobb sajátértékű faktorokat értelmezzük. Létezik azonban olyan javaslat is, mely szerint az összes varianciát 80%-ban magyarázó faktorokat célszerű kiszűrni. Gyakran olvashatunk olyan tanulmányokat, amelyek szerzői azt javasolják, hogy sok faktorra célszerű kiszámítani a sajátértékeket, és ezeket nagyság szerinti sorrendbe állítva a nagy „töréspontokig” célszerű kiszűrni a faktorokat.

A faktorok kiválasztását segíti a kommunalitás meghatározása is, amely megadja, hogy egy változó szórásának hányad részéhez járulnak hozzá a faktorok.

A faktoranalízis alkalmazásánál fontos a faktorsúly meghatározása. A faktorsúly megadja az egyes változók szerepét a kiszűrt faktorokban. Ebből megtudjuk a változók összefüggését az adott faktorokkal. Analóg a korrelációs együtthatóval, értékei és előjele is annak megfelelő. A faktorsúlyok előjele megmutatja, hogy a főfaktor alakulására az egyes mutatók milyen irányban hatnak. Azt is megvizsgálhatjuk, hogy a kapott első faktor milyen sztochasztikus kapcsolatban van az eredeti változókkal, és melyek azok az eredeti mutatók, amelyek legjobban jellemzik a területi egységeket. Sajnos az előjeleket nem veszik figyelembe, pedig a negatív faktorsúlyok alapján változtathatják meg egy faktor értelmezhetőségét.

A faktor érték (factor scores) az egyes vizsgált esetek „eredményei” a kiszűrt háttérváltozóban, faktorban az eredetileg mért változók alapján. Egyféle klasszifikációs, besorolási technika részeként is felhasználható.

A kérdés az, hogy mekkora faktorsúlyok tekinthetők lényegesnek? Erre nagyon nehéz válaszolni, mert egyértelmű szabály erre nincs. Általában támpontot ad a korrelációs együtthatók 5%-os szignifikancia szintje a (változók száma - 1) szabadságfok mellett. A gyakorlatban ez azt jelenti, hogy a 0,7 feletti faktorsú-

lyok mindig lényegesnek tekinthetők, de magas változós számok mellett a 0,5 feletti értékek sem hagyhatók teljesen figyelmen kívül. A többváltozós eljárásoknál, így a faktoranalízisnél is célszerű, ha a változók számánál legalább 50-nel nagyobb a vizsgált minta elemszáma. Az eredmények azonban ettől eltérő esetekben is lehetnek egyértelműek.

A faktorok száma általában lényegesen kisebb, mint a megfigyelt változók száma. Mivel a feltételek összességét a faktorok összessége jól visszatükrözi, ezért a vizsgált probléma lényegesen egyszerűbbé, áttekinthetőbbé válik. A faktorok, szemben az eredeti változókkal már előírhatóan páronként korrelálatlanok lehetnek. A faktorsémában is megfigyelt változókból következtetünk a megfigyelt változóra, a faktorok pedig mintegy a köztük lévő hatásokat közvetítik. Természetesen egy faktor nincs minden változóval feltétlenül kapcsolatban, ilyenkor a faktorsémában a megfelelő értékek - ezeket faktorsúlyoknak nevezzük - 0-val egyenlők. A faktorok értelmezése a felhasználó részéről elég nagy absztrakciós készséget, s előzetes ismereteket követel meg.

A faktoranalízis alaphipotézise szerint minden  $z_{ij}$  standardizált változóérték (ezzel a standardizálással hozzuk közös nevezőre az eltérő mértékegységű változókat) felírható fiktív, egyelőre önálló jelentéssel nem rendelkező változóértékeinek, az ún. faktorértékeknek lineáris kombinációjaként az alábbi formában:

$$z_{ij} = a_{i1}f_{1j} + a_{i2}f_{2j} + \dots + a_{ik}f_{kj} + a_i u_j$$

$$i = 1, \dots, N; j = 1, \dots, n$$

ahol  $f_{ij}$  jelenti az  $i$ -dik közös faktor  $j$ -dik megfigyelési objektumra vonatkozó értékét, az  $a_{ij}$  értékek a közös faktorsúlyok, az  $u_j$  értékek az egyedi és hibafaktorokat tartalmazzák, az  $a_i$  értékek pedig az egyedi faktorhoz rendelt faktorsúlyok.

A standardizált értékek segítségével jelentősen leegyszerűsödik a szorzatmomentum korrelációs mátrix:

$$R = \frac{1}{N} Z^* Z$$

ahol  $Z^*$  a  $Z$  transzponáltja.

Az  $R$  mátrix a következő tulajdonságokkal rendelkezik:

- kvadratikus, azaz  $(n \times n)$  típusú;
- szimmetrikus, azaz  $r_{jk} = r_{kj}$  minden  $j, k=1, \dots, n$  értékre;
- a főátlóban csak egyesek vannak.

Ez az  $R$  mátrix a faktoranalízis kiindulópontja. Az ismérvek közötti összefüggés a faktorsúlyok szorzatösszegeire vezethető vissza. Az  $R$  korrelációs mátrix vizsgálatából megállapítható, hogy a főátló csak egyesekből állhat. Ezek az önkorrelációk, amelyek a  $Z_j$  változók  $s_j^2$  teljes szórásnégyzetei, így az  $a_{jk}^2$  értékek-

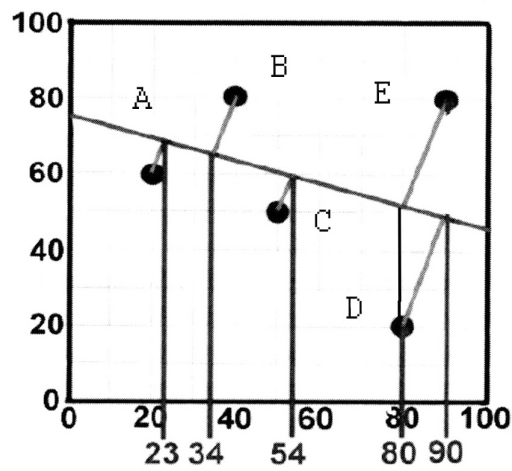
nek fontos szerepük van. Összegüket  $h_j^2$ -vel jelöljük és kommunalitásnak nevezük:

$$h_j^2 = a_{j1}^2 + a_{j2}^2 + \dots + a_{jm}^2 \quad j = 1, \dots, n$$

Ez a szórásnégyzetnek az a része, amely a közös faktorokra vezethető vissza. A kommunalításokat becsülni kell, ezek az eljárás végére alakulnak ki.

Az adatredukció egyik legegyszerűbb példáját (2 változó redukálása 1 változóra, pl. főkomponens-analízis) adja a következő diagram. Képzeljük el, hogy van 5 népcsoportunk, melyek génállományának sajátosságait (egy gén jelenlétének gyakoriságát az x és x kromoszómán) az alanti táblázat tartalmazza:

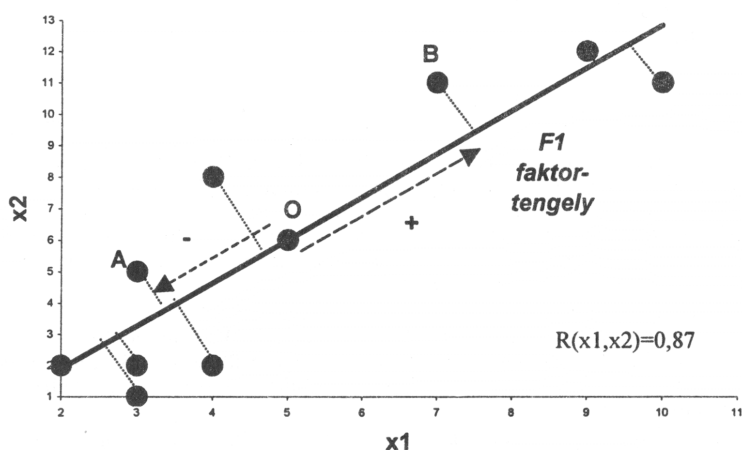
	X	Y
A	20	70
B	40	70
C	50	50
D	80	20
E	90	80



A, B, C, D, E pontok koordináta-rendszerben történő elhelyezése után – mivel a két változó között mérhető korreláció van – megrajzolható a regressziós egyenes. A pontokat az egyenesre vetítjük, a pontok egyenestől való távolságát felhasználva (mely egyben a két változó közötti korreláció erősségét, egyben a dimenziócsökkentés során bekövetkező az adatvesztés mértékét is jelzi), majd az így keletkezett pontokat az x tengelyre vetítjük. Ezzel a változók számát csökkentjük.

Egy másik módszert mutat be a következő ábra:

### A faktoranalízis logikája, a dimenziócsökkentés



Az adatredukció menete: A kétváltozós pontdiagramon kijelölhető a regressziós egyenes, melyre illeszkedik az adatok átlagértéke, mint O pont a koordináta-rendszerben. Az egyes pontok távolsága O-tól lesz a faktorérték (pozitív vagy negatív), így 2 számérték helyett immár egy jellemzi A és B pontot.

A helyesen kivitelezett faktoranalízisnek számos peremfeltétele van. Hasznos, ha a változók normáloszlást mutatnak. Az abszolút és a fajlagos adatok keverése nem ajánlott (pl. db és %, összeg és egy főre jutó érték)! Az azonos tartamra utaló változók szintén problémás (pl. GDP, GDP/fő, GDP%-os növekedése az előző évhez képest – ugyanakkor tény, hogy az össz-GDP két eltérő lakosságú ország esetében is lehet egyforma, azaz az egy főre jutó érték szerepeltetése a fajlagos teljesítőképesség, az össz-GDP-é pedig az ország teljesítőképessége miatt lehet fontos, így együttes használatuk igazolható).

A statikus és dinamikus adatok együttes szerepeltetése sem szerencsés (aktuális érték – érték %-os változása az előző évhez képest). Ez alól csak akkor tehetünk kivételt, ha a vizsgálatnak egyaránt célja mind az állapot, mind a fejlődési trend felmérése.

A jó faktoranalízis minél több változót igyekszik bevonni a vizsgálatba. Ugyanakkor egymásból lineárisan – pl.  $a+b=c$  képlettel – számolható változók szerepeltetése szintén nem ajánlott. Pl. a mezőgazdasági és ipari keresők arányának feltüntetése helyes, de a terciér szektorban dolgozókat már nem vonhatjuk be a vizsgálatba, hiszen az előző kettő ez utóbbi értékét is meghatározza, mivel 100%-ra egészítik ki egymást. Ugyanez a helyzet homok-iszap-agyagtartalom esetén is! Megfontolandó egymással szoros korrelációban lévő adatsorok alkalmazása is. Ezek ui. biztosan egy faktorba kerülnek: ha a faktorértékekkel számo-

lunk tovább, akkor ezek súlyozott hatása megjelenik az értékekben, tehát torzul az adatállomány, ha viszont csak az adott faktorra legerősebb korrelációt mutató egy, esetleg két változót használjuk fel a továbbiakban, eredeti értékekkel, az megoldja ezt a problémát. Az itt megfontolás tárgyává tett gondolatok a klaszteranalízisnél is érvényesek.

Faktoranalízist alkalmazhatunk statikus, ritkábban dinamikus vizsgálatoknál. Nagy idő-intervallumot felölelő vizsgálatoknál nem jelent feltétlenül problémát, ha egyik-másik adatsor nem ugyanabból az évből származik (kivéve, ha éppen akkor zajlott le valamilyen jelentős változás, mely nem teszi lehetővé az adatok összevetését). Nem meglepő módon az időbeli változások vizsgálata esetén előfordulhat, hogy megváltozik az adatszerkezet és a faktorok tartalma (más lesz pl. 1990-ben az első faktor tartalma és más lesz 2000-ben). Ezért dinamikus vizsgálatokhoz érdemben nem ad többletet a faktoranalízis (hacsak azt nem, hogy melyik változó besorolása a leginkább konstans és mely változók kerülnek át más faktorokba a korrelációs mátrix évenként eltérő koefficiensei miatt).

A faktorok értékeit a faktoranalízis elvégzése után összeadni stb. tilos, hiszen a faktorok lényege a dimenziótlanítás mellett éppen az, hogy egymással korrelálhatatlan változók keletkeznek! Ha a faktoranalízis egyetlen faktort eredményez, akkor rosszul válogattuk össze a változóinkat – pontosabban az általunk használt változók a valóságnak ugyanarra a részére utalnak (még ha ezt nem is gyanítottuk előzőleg), tehát helyettesíthetők.

A faktoranalízisnek több megoldási módszere van. Ennek a tantárgynak a keretében a feladatokat az SPSS szoftver segítségével oldjuk meg.

A faktoranalízis másik fő alkalmazási területét az ún. csoportosítási feladatok jelentik. Sajátos csoportosítási feladatként fogható fel az ún. tipizálási probléma is. Itt is területi egységeket jellemző mutatók sokasága áll rendelkezésre. A kérdés az, hogy mely mutatók vagy mutatócsoportok szerinti csoportosítás tükrözi legjobban a területi egységek alapvető jellegzetességeit. Ezek a mutatók természetesen más és más dimenzióban, de számszerű formában adóttak. A dimenzionális különbségek megszüntetése céljából a számítások során a mutatókat standardizált formában használjuk (mint erre már többször hivatkoztunk), ezek a faktormodell input változói. Az első faktort előállítva sokszor tulajdonképpen olyan komplex mutatószámot kapunk, mely lényegében a vizsgálandó fogalmat nagymértékben kifejezi, mivel a vele összefüggő mutatók információtartalmának zömét tartalmazza, s ennek a szintetikus mutatónak a faktorértékei (score-ok) alapján a rangsorolás elvégezhető. A faktorértékek alapján homogén csoportok alakíthatók ki, és így egy egydimenziós osztályozást hajtunk végre.

Az így kialakított csoportok térképen egyértelműen ábrázolhatók, s így szemléltethetők, sőt elemzést is végrehajthatunk ezek alapján.



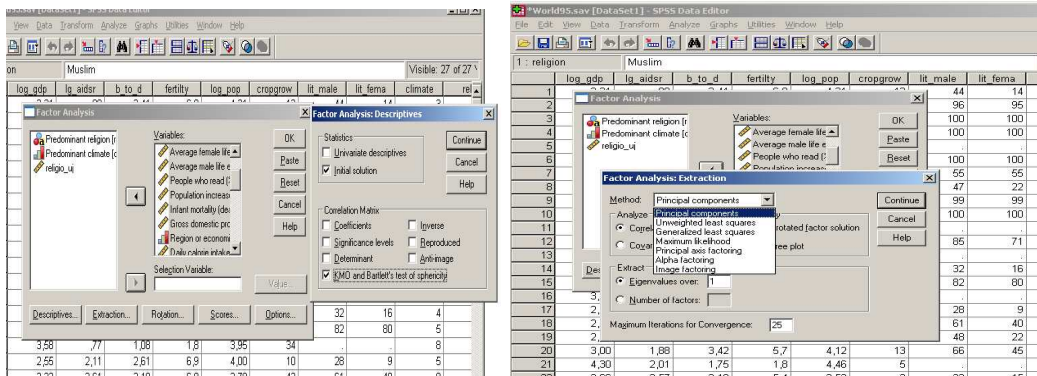
## 14.2. Faktoranalízis az SPSS-ben

SPSS-ben a faktoranalízist az **Analyze – Data reduction – Factor** paranccsal végezhetünk. A túl sok változó hátrányos lehet, ha kevés közöttük a független, különböző valóságtartamra utaló változó, de hátrányos lehet módszertani-ábrázolástechnikai szempontból akkor is, ha eseteinket csoportképzési szándékkal kívánjuk rendezni. Egy pontdiagram ugyanis kettő, maximum három tengellyel rendelkezik, ebből következően egyszerre 2–3 változó ábrázolható rajta.

A faktoranalízist alkalmazása során a különböző változókat „hasonlóságuk” alapján egy új változóba vonjuk össze. Az összevonás során kialakuló faktorok tartalmazhatnak egy vagy több változót, illetve egy változó is besorolásra kerülhet több faktorba (ez utóbbi nem túl szerencsés eset). Azt, hogy melyik változó melyik faktorba került besorolásra, azt az SPSS Rotated Component Matrix nevű, eredményeket tartalmazó táblázatában látható korrelációs értékek mutatják (az eredeti változó értékei páronként mennyire korrelálnak a kialakított faktorértékekkel). Egy változó mindig abba a faktorba kerül besorolásra, ahol a legnagyobb a korrelációs koefficiens mutatja (a pozitív és negatív értékek között nem teszünk különbséget). Előfordulhat, hogy egy faktorba csak egy változó kerül besorolásra, ebben az esetben e változó független változónak is tekinthető.

A faktoranalízis minőségi követelményeinek teljesülnie kell. Miután kiválogattuk a faktorba sorolni kívánt változókat a dimenziótlanításhoz (példánkban minden szám típusú változót faktorba sorolunk a **world95.sav** file-ból), a **Descriptives** gomb megnyomásával kérhetünk KMO-Bartlett tesztet, mely a faktoranalízis megbízhatóságát mutatja meg (0,7 feletti érték már jó). Kérhetünk korrelációs mátrixot is (ez hasznos a multikollinearitást mutató változók kiszűrésére). Az **Extraction** gomb megnyomásával egy újabb panelhez jutunk, ahol a faktorizálás módszere állítható be (javasolt a **principal component**, vagy a **maximum likelihood**). Itt állítható be, hogy a sajátérték =1 peremfeltételt választjuk, vagy pedig mi magunk állítjuk be a létrehozandó faktorok számát. Az előbbi esetben az eigenvalue<1 faktorok kidobásra kerülnek, mert kevesebbet magyaráznak az adatszerkezetből, mint bármelyik eredeti változó. Abban az esetben, ha pl. faktoraink száma 3, de csak a variancia 70%-át tartalmazzák, módosíthatunk a kialakítandó faktorok számán. A faktoranalízis ugyanis megbízható, ha a kialakított faktorok az eredeti variancia 80–85 %-át megőrzik. Értelemszerűen a két módszer ugyanazon adathalmaz esetén is eredményezhet eltérő eredményt, azaz a változók eltérő faktorokba kerülhetnek. Körültekintően kell tehát eljárni az eredmények felhasználásánál.

A maximális iterációs szám is átállítható, de esetek nagy részében már az automatikusan felajánlott 25-nél kevesebb iteráció esetén is megkapjuk a faktorokat. A **Rotation** gombnál a **Varimax** rotációt választjuk. A **Scores** paranccsal változóként elmenthetjük adattáblánkba a kapott faktorértékeket.



### KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0,786
Approx. Chi-Square		1895,550
Bartlett's Test of Sphericity	$\chi^2$	253
	Sig.	,000

A **KMO-Bartlett teszt** azt mutatja, hogy a faktoranalízis eredménye megbízható, a **Total Variance Explained** táblázat szerint 5 faktor mutatott 1-nél nagyobb sajátértéket, amelyek összesen az adatszerkezet varianciájának 82%-át tartalmazzák, tehát a veszteség nem jelentős, ebből az első faktor egymaga 43%-ot, a második 16%-ot magyaráz – kétváltozós diagram készítésekor tehát érdemes őket választani.

A **Rotated Component Matrix** tábla tartalmazza, hogy mely változók mennyire korrelálnak az adott faktoral. Példánkból megállapítható, hogy a nép-sűrűség gyakorlatilag használhatatlan változó (egy komponenssel sem mutatott jelentős korrelációt). Az 1. faktorba tartozik a legtöbb változó, az összes több mint fele, ezek tehát nem független változók. Az 5. faktorba csak a lélekszám (illetve annak 10-es alapú logaritmus) került, amit helyesebb kihagyni a vizsgálatból a fent elmondottak értelmében. 2. ua. tulajdonságra vonatkozó érték szerepeltetése nem szerencsés, miképp abszolút és fajlagos adatoké sem!). A lélekszám tehát független változónak tekinthető egy későbbi vizsgálat során. A GDP/fő a 2. faktort alkotja, s van olyan változó is, mely több helyre került, ezek használata szintén nem szerencsés. A 4. faktort az AIDS-ben szenvedők száma adja (ill. az értékek 10-es alapú logaritmus).

### Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squares and Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	10,014	43,538	43,538	10,014	43,538	43,538
2	3,624	15,756	59,294	3,624	15,756	59,294
3	2,534	11,015	70,309	2,534	11,015	70,309
4	1,561	6,786	77,096	1,561	6,786	77,096
5	1,338	5,816	82,912	1,338	5,816	82,912
6	,840	3,653	86,564			
7	,752	3,270	89,834			
8	,492	2,138	91,972			
9	,452	1,965	93,937			
10	,387	1,682	95,619			
11	,226	,982	96,601			
12	,198	,860	97,461			
13	,160	,696	98,157			
14	,110	,479	98,636			
15	,105	,457	99,093			
16	,056	,246	99,339			
17	,038	,167	99,506			
18	,033	,142	99,648			

Extraction Method: Principal Component Analysis.

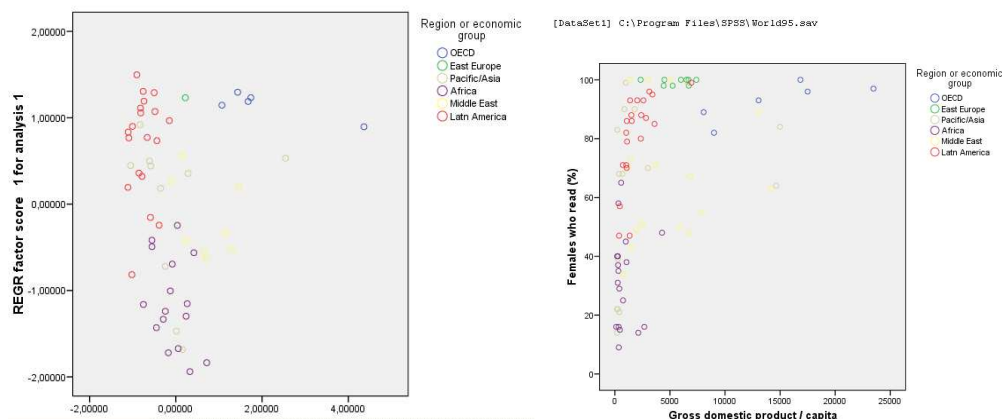
Ha ezek után arra vagyunk kíváncsiak, hogy elkülöníthetők-e csoportok pl. gazdasági fejlettség szerint (*region or economic group*), akkor vagy klaszteranalízist végzünk a faktorok/főkomponensek felhasználásával, vagy a **Graph - Legacy Dialog - Scatterplot** menüt választjuk ki, és kétváltozós pontdiagramon ábrázoljuk az eseteket. Az *x* és *y* tengelyre a két fő faktort visszük fel, a **Set markers by** ablakba pedig a *region or economic group* kerül. Ekkor kirajzolódnak az ország(csoport)ok elhelyezkedése a két fő faktor által létrehozott (matematikai és 2 dimenziós) térben, eltérő jellel jelölve a csoportokat

A két fő faktor értékei alapján készített pontdiagramon elkülöníthetők egyes gazdasági térségek (mások kevésbé). Ha a két **faktor** értékei helyett a két faktoralatt legnagyobb korrelációt mutató két változót ábrázoljuk, akkor némileg különbözik a pontok eloszlása, de hasonlóan elkülönülnek az egyes csoportok (ábra).

**Rotated Component Matrix(a)**

	Component				
	1	2	3	4	5
Population in thousands	,019	,001	-,138	-,153	<b>,800</b>
Number of people / sq. kilometer	,030	,306	-,385	-,341	-,342
People living in cities (%)	<b>,678</b>	,367	,278	-,124	-,281
Average female life expectancy	<b>,888</b>	,259	,160	-,284	-,017
Average male life expectancy	<b>,845</b>	,277	,199	-,341	,027
People who read (%)	<b>,964</b>	,043	-,027	,085	,005
Population increase (% per year)	<b>-,595</b>	-,057	<b>,724</b>	-,146	-,130
Infant mortality (deaths per 1000 live births)	<b>-,908</b>	-,245	-,110	,128	,045
Gross domestic product / capita	,435	<b>,842</b>	-,149	,139	-,055
Region or economic group	,063	<b>-,600</b>	<b>,555</b>	-,042	-,273
Daily calorie intake	<b>,595</b>	<b>,627</b>	,011	-,133	,032
Aids cases	,099	<b>,562</b>	,073	<b>,559</b>	,346
Birth rate per 1000 people	<b>-,871</b>	-,224	,314	,116	-,104
Death rate per 1000 people	<b>-,683</b>	-,062	-,447	,480	-,084
Number of aids cases / 100000 people	-,261	,056	-,106	<b>,865</b>	-,061
Log (base 10) of GDP_CAP	<b>,656</b>	<b>,630</b>	,079	-,025	-,249
Log (base 10) of AIDS_RT	-,027	,005	-,132	<b>,870</b>	-,292
Birth to death ratio	,065	-,044	<b>,837</b>	-,331	-,093
Fertility: average number of kids	<b>-,877</b>	-,113	,284	,124	-,126
Log (base 10) of Population	,083	,068	-,202	-,053	<b>,878</b>
crop grow	-,063	-,028	<b>-,572</b>	-,099	,368
Males who read (%)	<b>,914</b>	,035	-,048	,055	,072
Females who read (%)	<b>,964</b>	,012	-,042	,077	,012

Extraction Method: Principal Component Analysis.  
 Rotation Method: Varimax with Kaiser Normalization.  
 a. Rotation converged in 12 iterations.



**Példa:**

Példánkban a Bükkre jellemző morfológiai és litológiai paraméterek alkották az adatbázist. 25x25 m/pixel felbontást alkalmazva, ez mintegy 2 millió esetet eredményezett. Minden pixelhez hozzárendeltük a geológiát (kőzetkeménységet számként, ugyanis az SPSS szöveges változók esetében nem képes ezen műveletek végrehajtására, mi több, a szám típusú változókat arányskálaként kell értelmezni az SPSS-ben), lejtést, tszf.-i magasságot, kitettséget, fagyállóságot, erózióbázistól való távolságot, stb. Majd megvizsgáltuk, hogy ezen változók hogyan vonhatók össze. Nem okozott meglepetést, hogy a kőzetminőségre vonatkozó adatok (keménység, fagyállóság, kopásállóság, irodalmi denudációs érték, vízfeltevő-képesség) ugyanabba a faktorba nyertek besorolást – jóllehet fizikailag nem feltétlenül azonos tulajdonságra utalnak e változók).

Rotated Component Matrix/faktorok					
	1	2	3	4	5
Magasság Altitude	<b>0,555</b>	0,325	<b>0,577</b>	0,311	-0,017
UCS-keménység	<b>0,861</b>	0,192	0,249	0,079	-0,031
Fagyállóság Frost resist.	<b>0,918</b>	0,212	0,084	0,066	-0,007
Víz feltevő-kép. Porosity	<b>-0,787</b>	-0,058	-0,061	-0,043	-0,004
Kopásállóság Attrition resist	<b>0,859</b>	0,135	0,000	-0,054	0,031
Denudáció Denudation	<b>-0,919</b>	-0,205	-0,117	-0,088	0,018
Lejtés Slope	0,210	<b>0,834</b>	-0,004	-0,195	0,026
Costpush	0,236	<b>0,758</b>	0,137	0,317	0,008
Erozióbázis Distance.b.l.	0,099	0,019	<b>0,953</b>	-0,032	0,050
Letölés Runoff	0,059	0,012	0,016	<b>0,941</b>	0,004
Kitettség Aspect	-0,008	0,026	0,043	0,004	<b>0,998</b>
<b>KMO=0,77</b>	<b>45%</b>	<b>11%</b>	<b>9%</b>	<b>9%</b>	<b>8%</b>

Másik példánk a 2005-ös adatok alapján sorolja faktorokba a Köztes-Európa gazdasági fejlődésének vizsgálatánál alkalmazott változókat. Látható,

hogy az eredetileg több mint 30 változó viszonylag jól lett összeválogatva, mert 9 faktorba sorolhatók, igaz az 1. faktorba került közülük 14.

Faktoranalízissel kivont faktorok és részesedésük a teljes varianciából, ill. az egyes változók elhelyezkedése a faktorokon belül a 2005-ös adatsor esetében

Változók	Faktorok								
	1	2	3	4	5	6	7	8	9
FDI értéke	-0,058	0,021	-0,216	<b>-0,935</b>	-0,088	-0,07	0,013	-0,04	0,125
FDI/fő	-0,075	<b>0,705</b>	0,22	0,021	0,426	-0,05	-0,17	0,484	0,024
GDP 1990-hez képest	<b>0,418</b>	0,39	0,073	-0,38	-0,385	0,348	<b>0,474</b>	0,12	0,057
GDP előző évhez képest	-0,445	0,428	0,22	<b>0,609</b>	0,088	0,243	-0,26	0,09	0,199
állóeszközök	<b>-0,917</b>	0,014	0,08	0,296	0,068	0,031	0,176	-0,13	0,037
háztart. fogyasztás / fő	-0,657	-0,24	-0,284	<b>0,606</b>	-0,007	0,133	-0,15	0,051	0,022
keresked. cserearány	<b>0,824</b>	0,351	0,005	-0,332	-0,169	0,105	-0,06	-0,03	0,205
GDP/fő 92-höz képest	<b>0,834</b>	0,439	0,178	0,13	0,152	0,107	0	-0,12	0,109
Kereskedelmi egyenleg előző évhez képest	-0,181	-0,11	<b>-0,577</b>	-0,394	0,045	-0,57	-0,16	0,151	-0,18
fizetési mérleg hiány	<b>0,820</b>	0,113	-0,116	-0,333	-0,322	0,071	0,013	-0,26	-0,1
államházt. hiány	<b>-0,625</b>	0,387	-0,017	<b>0,275</b>	0,166	0,392	-0,09	-0,19	-0,4
áll.házt. hiány (előző év)	0,358	-0,46	0,36	<b>-0,536</b>	-0,121	-0,4	0,245	0,059	0,035
állattenyésztés	-0,173	-0,06	0,028	-0,039	-0,024	0,04	<b>0,914</b>	-0,16	0,22
növ.term.	-0,031	0,248	0,197	<b>0,688</b>	0,003	0,319	-0,04	0,429	0,278
gabona/ha	<b>0,801</b>	-0,05	0,129	-0,05	0,214	-0,41	0,318	-0,1	0,062
Ipari term./fő	-0,124	0,25	<b>0,717</b>	0,077	0,29	0,142	-0,44	0,324	0,06
Ipari term (előző évihez képest)	<b>0,966</b>	0,129	0,036	0,026	-0,013	0,075	0,028	-0,12	0,138
építőipar term.	-0,189	-0,12	0,092	0,012	-0,039	0,251	0,224	-0,15	<b>0,883</b>
CO <sub>2</sub> - kibocsátás/fő	0,211	<b>0,887</b>	0,128	-0,197	-0,061	-0,04	-0,01	0,229	-0,24
energiafelhasználás/ fő	0,391	<b>0,86</b>	0,229	0,071	0,067	-0,02	-0,1	-0,06	-0,09
hazasság-válás arány	-0,252	-0,41	-0,646	-0,221	-0,466	-0,22	-0,02	0,153	0,096
term.arányszám	-0,252	0,133	-0,029	0,183	0,103	-0,12	-0,18	<b>0,884</b>	-0,17
születés-halálozás	<b>0,618</b>	0,362	-0,348	-0,159	-0,541	-0,06	0,189	-0,04	-0,04
csecsemőhalandóság	<b>-0,765</b>	-0,28	-0,51	-0,155	0,175	-0,07	0,084	-0,02	0,014
keringési hal.	<b>-0,942</b>	-0,02	-0,047	-0,04	0,185	-0,19	-0,04	0,039	0,109
gazd.aktív lak.	0,245	<b>0,701</b>	-0,124	0,274	0,03	0,277	0,028	-0,34	0,405
munkanélküliség	-0,056	-0,15	0,105	-0,149	<b>-0,938</b>	-0,02	0,218	0,073	-0,03
tercier dolgozók	0,105	0,128	<b>0,937</b>	0,221	0,043	0,073	-0,04	0,082	0,161
reálbér	-0,071	0,138	<b>-0,819</b>	0,171	-0,023	0,145	-0,46	0,074	0,138
fogyasztói árindex	-0,67	-0,17	-0,546	0,101	<b>0,336</b>	-0,08	0,19	0,089	0,113
szgk/1000 fő	<b>0,707</b>	0,333	0,331	0,205	0,122	0,307	-0,15	-0,1	-0,31
k+f	<b>0,873</b>	0,177	0,089	0,101	0,269	-0,05	-0,16	0,031	-0,28
computer/1000	0,348	<b>0,668</b>	0,003	0,454	-0,072	0,309	0,182	0,303	-0,01
felsőoktatás/1000	0,098	0,031	0,098	0,176	0,103	<b>0,942</b>	0,023	-0,05	0,197
hulladék/fő	-0,107	0,004	0,213	-0,088	<b>0,853</b>	0,072	0,205	0,239	-0,11
autópálya/ter.	<b>0,803</b>	-0,16	0,021	0,406	0,178	-0,07	0,123	-0,16	-0,29
	<b>33%</b>	20%	10%	9%	7%	5%	4%	3%	2%

A keletkezett faktorokat felhasználhatjuk a klaszteranalízis-nél is (következő fejezet), amennyiben a sok azonos tartalomra utaló változó (gyakran eleinte éppen ezt nem tudjuk) helyett független változókkal akarunk dolgozni, vagy csak az esetek száma alá akarjuk csökkenteni a változók számát (a klaszterezés akkor ad megbízható eredményt, ha független változókkal végezzük el és a változók száma kevesebb, mint az esetek száma).

Ha a faktorok idegennek tűnnek, használhatjuk helyettük azt a változót, amely a legnagyobb korrelációs koefficiens-t mutatja az adott faktorial.

### 14.3. Főkomponens-analízis

A dimenzió csökkentés másik lehetséges módja a főkomponens analízis, amely a többváltozós matematikai statisztika egy széles körben elterjedt eljárása. Gyakran előfordul, hogy valamely változócsoporthoz mögött közös háttér, közös sok változó húzódik meg. Ezen háttérváltozó(k) kimutatásának eszköze lehet ez az eljárás. Egy másik lehetséges alkalmazás, amikor több dimenziós adatrendszerrel van dolgunk, és ebből mintegy kiemeljük a magyarázó, háttérváltozókat, lecsökkentve ezzel az adatrendszer dimenzió számát.

Abban az esetben, ha a faktoranalízis alkalmazási feltételei nem teljesülnek, akkor a főkomponens analízist alkalmazzuk. Ez a módszer a faktoranalízishez képest bizonyos szempontból speciálisabb, bizonyos szempontból viszont általánosabb. A főkomponens-elemzés általánosabb abban az értelemben, hogy itt a megfigyelési egységek száma lehet kisebb, mint a változók száma, azaz az eredeti standardizált változók korrelációs mátrixa lehet szinguláris, tehát nem teljes rangú mátrix is. Speciálisabb abban az értelemben, hogy itt nem feltételezünk specifikus főkomponenseket. Ez azt jelenti, hogy az alapmodellben a standardizált eredeti változók mátrixa a főkomponensek mátrixának és a főkomponens súlyok mátrixa transzponáltjának szorzatával egyenlő.

A számítások során a korrelációs mátrix főátlójában egyesek is állnak, tehát itt a kommunalitás értéke egy, mivel nem feltételezünk specifikus főkomponenseket, feltettük, hogy a változóknak nincs egyediségük. A számítások végén a főkomponensek száma megegyezik a változók számával, tehát nem annyira információsűrítésre, mint inkább a változók csoportosulásainak vizsgálatára használható.

A nagy főkomponens súlyok egy főkomponensbe tömörülése esetén a módszer dimenziócsökkentésre is alkalmas. A főkomponens-elemzés során kapott főkomponensek páronként korrelálatlanok. A feladatok megoldása alatt nem kell feltételezni az eredeti változók normális eloszlását, ezért szabadabban hasz-

nálható, mint a faktoranalízis, ugyanis bármilyen eloszlású adathalmaz vizsgálata alkalmas.

A főkomponens-analízissel csoportosíthatjuk az egyes változókat az egymás közötti korrelációjuk alapján. Így felismerhetők, hogy mely változók tartoznak össze, hány ilyen csoport van és csoporton belül milyen irányú és mennyire szoros a változók összefüggése. Azonban választ nemcsak változópáronként kaphatunk, hanem egyszerre áttekinthetjük az összes változó egymás közötti korrelációs rendszerét. A háttérváltozók a főkomponens analízis keretén belül a faktorok. Az analízis célja tehát, hogy a komplexitás magas fokát, amelyet a változók sokasága jelent, kezelhetővé és könnyebben értelmezhetővé tegye, a változókat a lehető legkevesebb faktorról magyarázza. Általában sem a faktorok tulajdonsága, sem azok száma előre nem ismert. Egy jó és sikeres faktoranalízis azzal jellemezhető, hogy a mintában levő sok változót csak néhány faktor reprezentál, ugyanis semmit nem nyernénk, ha közel annyi faktorra lenne szükség, mint amennyi a változók száma, hogy az összefüggéseket jellemezhessük.

### ***Dimenziószám csökkentés főkomponens analízissel***

Legyen  $p$  számú megfigyelési egységünk, amelyek egyenként  $n$  számú adatot tartalmaznak ( $p$  számú megfigyelési vektorunk van).

$x^1$	$x^2$	.....	$x^p$
$x_1^1$	$x_1^2$		$x_1^p$
$x_2^1$	$x_2^2$		$x_2^p$
...	...	.....	...
$x_n^1$	$x_n^2$		$x_n^p$

Tekintsük az  $x_j$  vektorokat valószínűségi változóknak, a vektorok elemeit a valószínűségi változók realizációinak. Standardizáljuk a változókat:

$$\overline{x_i^j} = \frac{x_i^j - \overline{x^j}}{s^j}$$

ahol  $x_j$  a  $j$ -edik vektor elemeinek átlaga (a várható érték becslése), és  $s_j$  az empirikus szórása. Így tehát 0 várható értékűvé, és 1 szórásúvá tettük a valószínűségi változóinkat.

Ezek után számítsuk ki az adatrendszerünk korrelációs mátrixát:



$$\underline{R} = \begin{pmatrix} r_{11} & r_{21} & r_{31} & \dots & r_{1p} \\ r_{21} & r_{22} & r_{32} & \dots & r_{2p} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ r_{p1} & r_{p2} & r_{p3} & \dots & r_{pp} \end{pmatrix}$$

ahol  $r_{ij}$  az  $i$  és  $j$ -edik megfigyelési egységek korrelációs együtthatója.

Határozzuk meg a korrelációs mátrix sajátértékeit és sajátvektorait, vagyis oldjuk meg a következő sajátérték egyenletet:

$$\underline{R} v = \lambda v$$

A kapott sajátértékek  $\lambda_1, \lambda_2, \dots, \lambda_p$ , a sajátvektorok  $v_1, v_2, \dots, v_p$

Számítsuk ki a főkomponenseket a következő módon, legyen a  $j$ -edik főkomponens a következő:

$$C_i^j = \sum_p x_i^p v_p^j$$

ahol  $i=1, n$  és  $j=1, p$ .

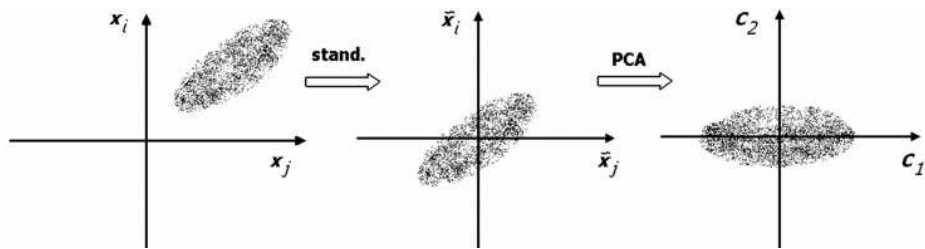
A főkomponensek ortogonális rendszert alkotnak, vagyis korrelálatlanok, azaz korrelációs mátrixuk

$$\underline{R}_C = \begin{pmatrix} \lambda_1 & & & & 0 \\ & \lambda_2 & & & \\ & & \dots & & \\ & & & \dots & \\ 0 & & & & \lambda_p \end{pmatrix}$$

A  $\underline{R}_C$  fontos tulajdonsága, hogy a főkomponensek és a standardizált változók összvarianciája azonos:

$$\sum_{j=1}^p \lambda_j = \sum_{i=1}^p s_i^{-2} = \sum_{j=1}^p s_j^2 = p$$

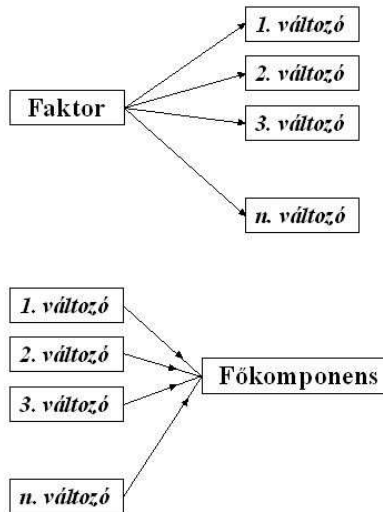
Amint látható, a főkomponensek kiszámításával nagymértékben átrendeztük a varianciákat, mivel, ha ez lehetséges volt, összevontuk őket az első (néhány) főkomponensben. Az eljárás főbb mozzanatainak geometria jelentése a következő:



Abban az esetben, ha például az első főkomponens képes magába sűríteni a megfigyelési egységek variáciáinak nagy részét, akkor megtehetjük, hogy az egész adatrendszert csak az első főkomponensével helyettesítjük. Ezzel végül is nem követünk el túl nagy hibát, viszont jelentős mértékben csökkentjük az adatrendszer dimenziószámát, ezzel adatszámát, és így meggyorsítottuk, megkönnyítettük egy soron következő eljárás, például a klaszter analízis működését.

A faktoranalízisnek és a főkomponens-analízisnek egyaránt elsődleges célja az adatsökkentés. Gyakran nagyszámú változóval dolgozunk, amelyek egymással korrelálnak. Ezek számát a kezelhetőség érdekében csökkenteni kell. Az elemzés során az egymással kölcsönösen összefüggő változók közötti kapcsolatokat vizsgálunk, és ezeket néhány magyarázó faktorral, főkomponenssel jelenítjük meg.

A faktoranalízis és a főkomponens analízis hasonlít egymásra, de az okság iránya fordított a két eljárásnál: a faktoranalízisnél az okság a globális értékek, vagyis faktorok felől mutat a változók felé, míg a főkomponens analízisnél az okság a változóktól a globális értékek, vagyis a főkomponens felé mutat.



### Feladatok

1. A world95.sav file felhasználásával alkosson faktorokat a változókból, állapítsa meg, melyek a főkomponensek a faktoron belül, majd a két fő faktor segítségével készítsen kétváltozós pontdiagramot, ahol külön jellel jelenik meg minden egyes vallás/gazdasági régió. Elkülöníthetők-e a csoportok a faktorok értékei alapján, vagy összemosódnak?
2. A meteorológiai állomásokon mért adatok alapján állapítsa meg, hogy redukálható-e a változók száma!

Állomás	Csapadék (mm)	Évi átlag-hőmérséklet (°C)	Potenciális párolgás (mm)	Fagymentes napok száma	Napsütéses órák száma
1	3250	25	2530	354	2850
2	1520	23	1560	298	2209
3	325	3	120	50	1650
4	654	14	450	258	1899
5	120	20	400	360	3014
6	985	9	875	185	2005
7	2001	15	1500	299	2006
8	1253	8	863	315	1750

3. A következő oldalakon látható változók értékei alapján készítsen korrelációs mártixot a változók és az országok között (normalitásvizsgálat, szükség szerint  $\lg(x+1)$  alkalmazása).
4. Az egy főre jutó direkt adó milyen kapcsolatot mutat az egy főre jutó költségvetéssel? És az egy főre jutó államadósság és az egy főre jutó költségvetés értéke milyen viszonyban van egymással? Hogyan befolyásolja ezt a viszonyt az egy főre jutó indirekt adó/ az egy főre jutó kivitel/ az egy főre jutó hadikiadás?
5. Fennáll-e a multikollinearitás veszélye az alanti táblázat adatainál? Válasszon ki olyan változókat melyek szerepeltethetők egy regresszió-analízisnél! Az itt látható adatsor mely feltételek között alkalmas faktoranalízisre? Hány faktorba sorolhatók a változók?
6. Mely változók határozzák meg/befolyásolják/tükrözik leginkább az államadósság értékét? És az oktatásra ill. hadsereg fordított fajlagos (egy főre jutó) költségeket?
7. Mely változók *nem* határozzák meg/befolyásolják/tükrözik az egy főre jutó hajótérfogatot? És az államadósság törlesztését?
8. Érdemes-e bevonnni egy klaszterképzést célzó vizsgálatba ennyi (hasonló tartamra utaló abszolút ill. fajlagos) változót (függetlenség kritériuma, sokaság)? Mely változókat érdemes kiválasztani?

1898	Állam- adósság (millió K)	állam- adósság /fő	Államadósság törlesztése (millió K)	törlesztés /fő	Államadósság /kölségvetés	adósságtörlesztés/ kölségvetés (%)
Ausztria	8500	354	341	14	6,07	24
Belgium	2330	364	114	18	6,01	29
Bulgária	169	51	18	5	2,04	21
Dánia	260	124	33	16	2,83	35
Franciao.	31000	816	1250	33	9,23	37
Görögo.	820	342	22	9	8,63	23
Hollandia	2400	490	182	37	8,28	62
Magyaro.	4500	257	64	4	4,52	6
Nagy-Brit.	16000	410	625	16	6,40	25
Németo.	15000	288	90	2	9,38	5
Norvégia	210	105	10	5	2,21	10
Olaszo.	13000	417	810	26	7,74	48
Oroszo.	15000	142	670	6	2,50	11
Portugália	3500	700	90	18	12,73	32
Románia	1120	193	80	14	5,21	37
Spanyolo.	6400	366	324	19	7,25	37
Svájc	340	117	4	1	2,00	2
Szerbia	408	177	18	8	6,48	18
Svédó.	400	82	33	7	2,35	19
USA	5000	81	200	3	2,86	11
Japán	2200	50	160	4	1,57	11
Töröko.	4150	180	175	8	9,22	39
Kína	1650	4	9	0	3,06	2
Monarchia	13000	283	405	9	4,81	15

1898	Kölség- vetés Millió K	kölség- vetés /fő	direkt adó millió K	Direkt adó %	Direkt adó/fő	Indirekt adó* millió K	Indirekt adó %	Indirekt adó/fő	direkt/ indirekt adó	Kölség- vetés /kivitel
Ausztria	1400	58	220	16	9	725	52	30	0,30	0,93
Belgium	388	61	55	14	9	140	36	22	0,39	0,28
Bosznia	36	24	9	25	6	10	28	7	0,90	0,40
Bulgária	83	25	33	40	10	35	42	11	0,94	0,77
Dánia	92	44	13	14	6	61	66	29	0,21	0,25
Franciao.	3360	88	521	16	14	2000	60	53	0,26	0,75
Görögo.	95	40	20	21	8	51	54	21	0,39	1,36
Hollandia	290	59	66	23	13	150	52	31	0,44	0,10
Magyaro.	996	57	185	19	11	360	36	21	0,51	1,00
Nagy-Brit.	2500	64	408	16	10	1680	67	43	0,24	0,35
Németo.	1600	31	?	?	?	870	54	17	?	0,36
Norvégia	95	48	5	5	3	50	53	25	0,10	0,53

Olaszo.	1680	54	480	29	15	880	52	28	0,55	1,41
Oroszo.	6000	57	300	5	3	1800	30	17	0,17	2,31
Portugália	275	55	60	22	12	150	55	30	0,40	1,83
Románia	215	37	33	15	6	61	28	11	0,54	0,66
Spanyolo.	883	50	300	34	17	307	35	18	0,98	0,96
Svájc	170	59	1	1	0	46	27	16	0,02	0,23
Szerbia	63	27	21	33	9	5	8	2	4,20	1,19
Svédó.	170	35	28	16	6	86	51	18	0,33	0,40
USA	1750	28	1	0	0	1600	91	26	0,00	0,35
Japán	1400	32	125	9	3	250	18	6	0,50	1,56
Töröko.	450	20	200	44	9	180	40	8	1,11	1,25
Kína	540	1	360	67	1	150	28	0	2,40	0,68
Monarchia	2700	59	410	15	9	1100	41	24	0,37	1,04

1898	Kivitel Mó K	Behoz mó K	kivitel/ behoz.	kivitel /fő	Beho- zatal /fő	Mezg mó fő	Mezg %	Ipar, bánya mó fő	Ipar%	Keres- kedő mó fő	Keres- kedő %
Ausztria	1500	1400	1,07	63	58	13,5	56,2	6,10	25,42	2,20	9,17
Belgium	1400	1700	0,82	219	266	0,50	7,81	1,10	17,19	0,30	4,69
Bosznia	90	100	0,90	60	67	1,20	80,0	0,05	3,33	0,05	3,33
Bulgária	108	76	1,42	33	23	3,0	90,9	0,10	3,03	0,10	3,03
Dánia	370	504	0,73	176	240	1,00	47,6	0,53	25,24	0,17	8,10
Franciaó.	4500	4900	0,92	118	129	8,20	21,5	4,40	11,58	2,10	5,53
Görögo.	70	113	0,62	29	47	0,44	18,5	0,10	4,17	0,12	5,00
Hollandia	2800	3500	0,80	571	714	0,54	11,0	0,53	10,82	0,26	5,31
Magyaró.	1000	2000	0,50	57	114	6,20	35,4	0,96	5,49	0,20	1,14
Nagy-Brit.	7200	10720	0,67	185	275	2,50	6,4	9,00	23,08	1,60	4,10
Németó.	4500	5500	0,82	87	106	18,0	34,6	20,0	38,46	6,00	11,54
Norvégia	180	330	0,55	90	165	0,75	37,5	0,35	17,50	0,20	10,00
Olaszo.	1190	1070	1,11	38	34	9,00	28,8	4,10	13,14	0,60	1,92
Oroszo.	2600	2000	1,30	25	19						
Portugália	150	205	0,73	30	41						
Románia	324	337	0,96	56	58	1,50	25,8	0,45	7,76	0,14	2,41
Spanyolo.	924	793	1,17	53	45	4,80	27,4	1,00	5,71	0,30	1,71
Svájc	730	1060	0,69	252	366	1,10	37,9	1,00	34,48	0,20	6,90
Szerbia	53	33	1,61	23	14	0,70	30,4	0,40	17,39	0,40	17,39
Svédó.	420	450	0,93	86	92	1,20	24,4	0,70	14,29	0,31	6,33
USA	5000	4000	1,25	81	65	11,0	17,7	6,00	9,68	7,00	11,29
Japán	900	620	1,45	20	14	17,0	38,6	2,70	6,14	4,60	10,45
Töröko.	360	340	1,06	16	15						
Kína	800	1200	0,67	2	3						
Monarchia	2600	2500	1,04	57	54	21,0	45,6	7,10	15,43	2,40	5,22

1898	Hajó 1000t	Hajó t/fő	Vasút 1000 km	vas- út/ter	Hadi- teng. Mó K	Hadi- teng %	Hadi- teng /fő	Ókta- tás Mó K	Okta- tás %	Okta- tás /fő	Hadügy móK	Hadügy %	hadügy/ fő
Ausztria	180	8	16,00	0,05	29	2,07	1,21	58	4,14	2,42	330	23,57	13,75
Belgium	85	13	5,00	0,17	1	0,26	0,16	26	6,70	4,06	48	12,37	7,50
Bulgária			0,96	0,01	1	1,20	0,30	9	10,84	2,73	22	26,51	6,67
Dánia	346	165	2,30	0,06	9	9,89	4,33	5	5,76	2,52	13	14,13	6,19
Franciao.	894	24	37,00	0,07	284	8,45	7,47	213	6,34	5,61	629	18,72	16,55
Görögo.	588	245	0,90	0,01	5	5,26	2,08	6	5,89	2,33	18	18,95	7,50
Hollandia	835	170	2,60	0,08	33	11,38	6,73	30	10,34	6,12	46	15,86	9,39
Magyaro.	12	1	14,00	0,04	1	0,10	0,06	26	2,61	1,49	32	3,21	1,83
Nagy-Brit.	10500	269	34,00	0,11	550	22,00	14,10	280	11,20	7,18	450	18,00	11,54
Németo.	1500	29	47,00	0,09	72	4,50	1,38	300	20,00	5,9	590	36,88	11,35
Norvégia	330	165	1,70	0,01	10	10,53	5,00	11	11,58	5,50	9	9,47	4,50
Olaszo.	776	25	15,70	0,05	96	5,71	3,08	42	2,50	1,35	235	13,99	7,53
Oroszo.	400	4	40,00	0,01	270	4,50	2,55	104	1,73	0,98	1100	18,33	10,38
Portugália	90	18	2,30	0,03	20	7,27	4,00	5	1,82	1,00	30	10,91	6,00
Románia	61	11	2,80	0,02	1	0,47	0,17	27	12,56	4,66	45	20,93	7,76
Spanyolo.	486	28	12,30	0,02	28	3,17	1,60	48	5,44	2,74	142	16,08	8,11
Svájc			3,70	0,09				60	35,29	20,69	25	14,71	8,62
Szerbia			0,57	0,01				5	7,62	2,09	14	22,22	6,09
Svédó.	483	99	9,80	0,02	20	11,76	4,08	19	11,18	3,88	25	14,71	5,10
USA	4500	73	293,0	0,03	200	11,43	3,23				250	14,29	4,03
Japán	255	6	3,70	0,01	50	3,57	1,14	11	0,79	0,25	160	11,43	3,64
Töröko.	249	11	4,50	0,00	13	2,89	0,57	4	0,89	0,17	125	27,78	5,43
Kína	616	2	0,90	0,00	20	3,70	0,05	5	0,93	0,01	100	18,52	0,25
Monarchia	192	4	31,00	0,05	30	1,11	0,65	84	3,11	1,83	352	13,04	7,65

## 15. Klaszteranalízis

A földtudomány különböző területein gyakran nagyméretű adatbázisok elemzésével, értékelésével foglalkozunk. Ezekben sokszor nehézséget jelent az eligazodás. Az adatok tulajdonképpen különféle információt jelentenek számunkra, és ezek alapján lehetőségünk nyílik csoportosításukra, vagyis az adatbázis elemeinek különféle osztályozására. Az osztályba sorolásnál az egyes csoportok kialakításánál tulajdonképpen előre meghatározzuk (definiáljuk) azok minőségi paramétereit. A klaszterezés is tulajdonképpen egyfajta osztályozás azzal a különbséggel, hogy itt előre nem tudjuk milyen minőségi osztályba fognak tartozni az előálló csoportok, sőt ezek határai sem ismertek. A klaszteranalízis tehát az összefüggések halmazát vizsgálja, és nem tesz különbséget függő és független változó között, hanem a változók halmazán belüli kölcsönös összefüggéseket vizsgálja. A klaszteranalízis fő célja, hogy a kiválasztott változók alapján az adataink homogén csoportokba rendeződjenek. Az egyes csoportba tartozó elemek hasonlítanak egymásra, és különböznek más csoportok tagjaitól.

A klaszteranalízis lényege az adatok (esetek) besorolása csoportokba a jellemző tulajdonságok (változók) alapján. Fontos peremfeltétel, hogy a vizsgált változók száma kisebb legyen, mint az esetek száma.

Elsősorban akkor alkalmazzuk, ha

- jól felépített adatbázissal rendelkezünk, ahol minden egyes esethez több tulajdonság (változó) tartozik, s az egyes esetek összetartozása nem dönthető el csupán egy-egy változó értékei alapján. (pl. egy mélyfúrásban egy-egy réteg jellemezhető színnel, szemeloszlással, fossziliatartalommal, karottázs-adatokkal, elemösszetétellel, mechanikai tulajdonságokkal, stb. Ennek alapján a rétegek összetartozása minősíthető külön-külön (is) minden egyes változó esetében, de egyáltalán nem szükségszerű, hogy egyformán csoportosítsuk a rétegeket, vagy húzzuk meg a réteghatárokat minden változó esetében. Ekkor klaszteranalízist alkalmazva minden változó figyelembe vételével azonosíthatók a hasonló és egymástól különböző rétegek.)
- térbeli (akár foltszerű, akár pontszerű) elterjedést mutató egységeket kívánunk csoportokba rendezni (pl. területek fejlettségét szeretnénk jellemezni sok változó – GDP/fő, munkanélküliség, terciér szektorban dolgozók aránya, egyetemet végzettek aránya, egy főre eső nettó jövedelem, stb. – alapján)
- tematikus térkép szerkesztésénél az összes változó felhasználásával egy térképen ábrázoljuk csoportokba sorolt eseteket, s vizsgálhatjuk a térbeli elterjedés sajátosságait. (pl. az azonos klaszterbe tartozó területek határosak-e egymással vagy elkülönülnek, ennek mi az oka. Pl. ha vízmintáink vannak különböző helyekről, s mindegyikből van adatunk pH-ra, ortofoszfát-tartalomra, nitrát-tartalomra és szervesanyagmennyiségre, ezáltal a mintavételi pontok/vízfolyásszakaszok osztályozhatók, stb. Ha rendelkezésre állnak meteorológiai adatok (csapadék, napsütéses órák száma, potenciális párolgás, fagymentes napok száma, stb.), akkor lehetőség van a klímátípusok vizsgálatára, elkülönítésére is.

- túl sok kategóriát tartalmaz a megjelenített tulajdonság és ezeket kívánjuk csoportokba összevonni. (pl. nem ellenőrzött osztályba sorolás műholdfelvételek kiértékelésekor. Ebben az esetben a megjelenített kompozit kép akár 255 színt is mutathat, azaz a pl. vörössel jellemezhető erdőterületek maguk is több színárnyalattal rendelkeznek. A klaszterezés során a színárnyalatok száma – a kompozitban felhasznált csatornák reflektancia-értékeinek felhasználásával – csökkenthető, így elérhető, hogy az erdőterületek ne vöröses, hanem vörös színben jelenjenek meg.)
- ún. idősor analízist választunk, ekkor a változók évszámok lesznek, s az eseteket aszerint soroljuk klaszterekbe, hogy milyen az időbeli változás lefutása, trendje. (pl. az export és import trendjének változása az egyes országoknál. A hasonló trendet mutató esetek kerülnek azonos klaszterbe. A hasonlóság jelenthet volumet – értéknagyságot vagy jelenthet trendet is, ez eltérő standardizálási technikát igényel)

A klaszteranalízis szinte bármilyen területen felhasználható, ahol egy adatponthoz – akár mutat felszíni elterjedést, akár nem – több változó tartozik. Használható tehát a geológiában, meteorológiában, társadalomföldrajzban, természetföldrajzban, de bármelyik természettudományban is.

Azt is hangsúlyoznunk kell, hogy a klaszterezés során a hasonlóságot és különbséget vizsgáljuk, azaz a kapott eredmény a minőségi különbségek megállapítását önmagában nem teszi lehetővé. Az adatok ismerete nélkül nem mondhatjuk, hogy az egyik klaszterbe tartozó esetek „jobbak”, vagy kevésbé „jobbak”. Ahhoz az eredeti adatokhoz tartozó változók értékeit is meg kell nézni, s csak ez után állítható fel a klaszterek között minőségi rangsor. Sok esetben ennek nincs is értelme: pl. a már említett fűrés rétegeinek minősítése esetében a hasonlóság és különbség megállapításának sokkal nagyobb jelentősége van, mint annak, hogy melyik mutat nagyobb értékeket.

### **15.1. A klaszterelemzés fajtái**

A klaszterelemzésnél alapvető feladat azoknak a változóknak a megtalálása, amelyek a csoportok közötti különbséget okozzák. A változók kiválasztása elméleti megfontolások, korábbi kutatások, stb. figyelembe vételével történhet.

Az adatbázisunkban a korrekt osztályozás csakis klaszterezési algoritmusok révén valósítható meg. A klaszterelemzés algoritmusai lehet hierarchikus vagy nem hierarchikus. A hierarchikus algoritmus az új klasztereket az előzőleg kialakított klaszterek alapján keresi meg, a nem hierarchikus algoritmus pedig egyszerre határozza meg az összes klasztert.



## *Hierarchikus eljárások*

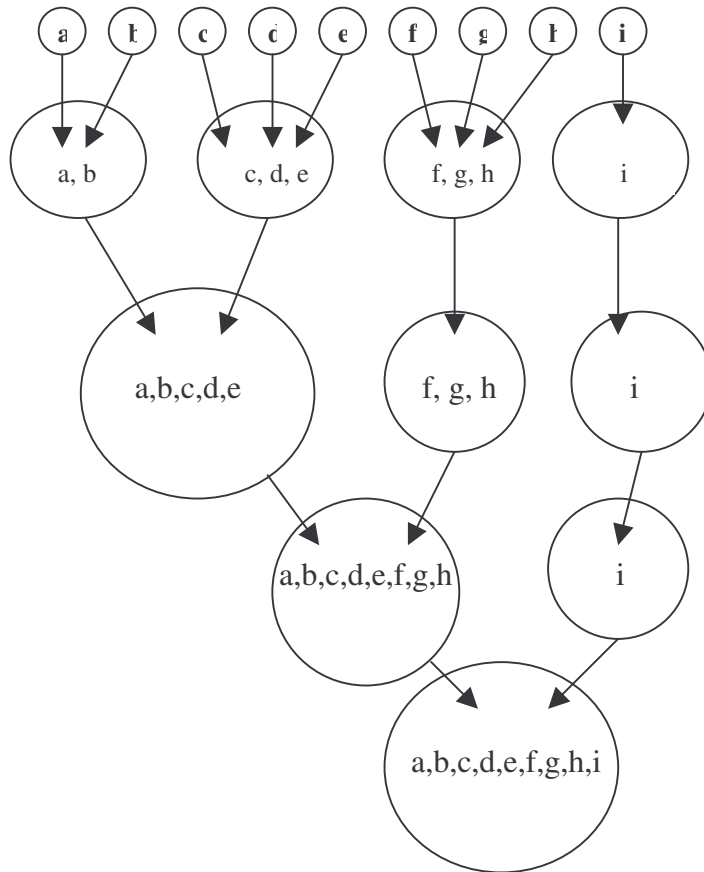
A hierarchikus klaszterező eljárásokban az adatokat hierarchikus adatszerkezetbe (fába, dendrogram) rendezzük. Az egyes adatok a „fa leveleiben” helyezkednek el. A fa minden belső elágazási pontja egy klaszternek felel meg, és azokat a pontokat tartalmazza, amelyek a fában alatta találhatók.

A dendrogram ágszerkezete segít annak eldöntésében, hogy hány klaszterbe kívánjuk eseteinket besorolni – ez ugyanis szintén módszertani kérdés: a klaszterek száma 1-től  $n$ -ig terjedhet. Az első esetben minden eset ugyanabba a klaszterbe kerül, ez látható a jobb oldalon a „fa törzseként” megjelenítve, míg az utóbbi esetben minden eset ( $n$  db), önálló klasztert képvisel, ezek láthatók a bal oldalon kis „ágakként”. Nyilvánvaló, hogy az általunk használni érdemes megoldás valahol a kettő között helyezkedik el. Az ágak hosszúsága a csoportok közötti relatív távolságokat reprezentálja. Ha a dendrogramot egy vonallal tetszőleges helyen elmetsszük, akkor a metszéspontok száma adja meg az éppen aktuális klaszterek számát és leolvasható az is, hogy mely esetek tartoznak hozzá. Ezt követően a dendrogramról leolvasott adatokból előállítható a foltterkép.

Két alapvető hierarchikus eljárás létezik: az egyik a **felhalmozó**, a másik a **lebontó**. A felhalmozó eljárásban kezdetben minden adatelem egy klaszter, majd a legközelebbi klasztereket egyesíti az algoritmus, és a hierarchiában egy szinttel feljebb új klasztert alakít ki. Először nézzünk egy példát a felhalmozó, tehát az adatelemek összevonásán alapuló klaszterezésre. Tekintsük át a településeket tartalmazó adatbázist a népességszám alapján.

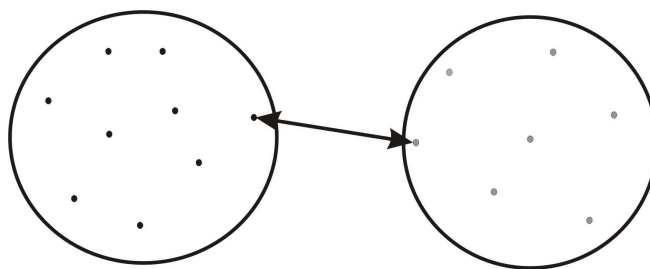
Jelöljük **a**-val a tanyákat, **b**-vel a bokortanyákat, **c**-vel a kis községeket ( $< 2000$  fő), **d**-vel a közepes méretű községeket ( $2000 < d < 5000$  fő), **e**-vel az 5000 főnél nagyobb lakosságszámú községeket, **f**-vel a kisvárosokat ( $< 20000$  fő), **g**-vel a közepes lakosságszámú városokat ( $20000 < g < 80000$  fő), **h**-vel a nagy lakosságszámú városokat ( $80000 < h < 220000$ ) és **i**-vel a fővárost. Ebben a példában kilenc elem van: {a}, {b}, {c}, {d}, {e}, {f}, {g}, {h} és {i}. Az alábbi dendrogramon látható, hogy az algoritmus hogyan kapcsolja össze az elemeket.

Az ábráról láthatjuk, hogy az algoritmus először összekapcsolta a legközelebbi elemeket, és ekkor a következő klasztereink lesznek: {a, b}, {c, d, e}, {f, g, h}, valamint az {i}. Tehát a tanyák, községek, városok és a főváros kerülnek egy-egy klaszterbe. Ezután a tanyák klasztert a legközelebbi, leginkább hasonló elemmel, a községekkel kapcsolja össze. Így már csak három klaszterünk lesz: a kis lélekszámú települések, városok és a főváros. A következő lépésben a kis települések és a városok klasztere lesz a legközelebb egymáshoz, ezért ezekből jön létre egy új, nagyobb klaszter. Most már csak két klaszterünk van, az egyik a főváros, a másik pedig az összes települést tartalmazza. Az összevonáson alapuló eljárás végeredményeként pedig egyetlen nagy klaszter jön létre, mely tartalmazza az összes településtípust: {a, b, c, d, e, f, g, h, i}.

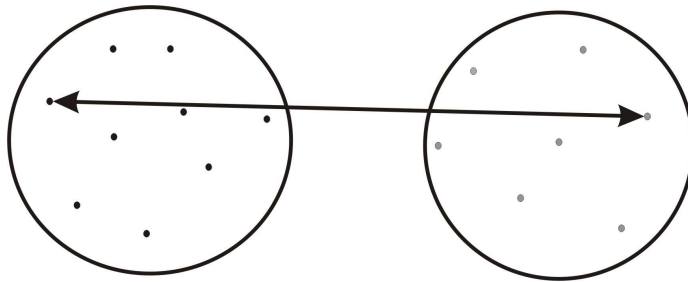


A felhalmozó (összevonáson alapuló) klaszterelemzésen belül különféle csoportosítási módszereket szoktak alkalmazni.

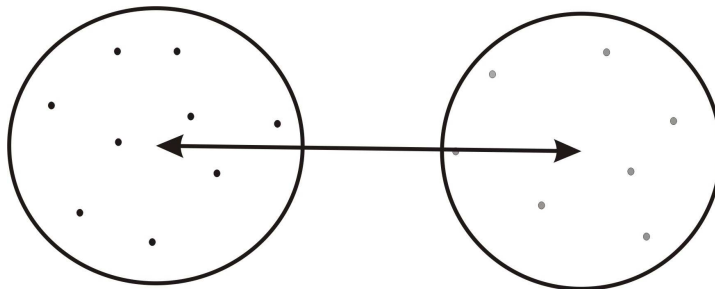
- Az **egyszerű láncmódszer** a legközelebbi szomszéd (nearest neighbor) elvén alapuló eljárás, amely a két klaszter közötti távolságot a két legközelebbi elem távolsága alapján számolja ki.



- A **teljes láncmódszer** a legtávolabbi szomszéd (furthest neighbor) elvén alapuló eljárás, amely a két klaszter közötti távolságot a két legtávolabbi elem távolsága alapján határozza meg.



- Az **átlagos láncmódszer** (Average linkage clustering) alkalmazásakor a két klaszter távolságát az összes elem páronkénti távolságának átlaga alapján határozzuk meg, ahol a pár egyik eleme az egyik klaszterbe, a másik eleme pedig a másik klaszterbe tartozik. Ennek a módszernek az az előnye az előző kettővel szemben, hogy nemcsak a legkisebb és legnagyobb távolságot használjuk fel.
- A **varianciamódszer** (Ward-féle eljárás) az egyik leggyakrabban alkalmazott módszer. Ennél a klaszteren belüli szórásnégyzet minimalizálásával állítunk elő klasztereket. Első lépésként minden egyes klaszterre kiszámoljuk az összes változó átlagát. Ezután a klaszterátlaghoz képest kiszámítjuk a négyzetes euklideszi távolságot minden elemre. Minden lépésnél azt a két klasztert vonjuk össze, amely a klaszteren belüli szórásnégyzetet a legkevesbé növeli.
- A **centroidmódszer** lényege, hogy a két klaszter közötti távolságot a klaszterek középpontja (centroidja) alapján határozzuk meg. A klaszterek centroidját a klaszterben bennfoglalt összes pont átlaga adja. A legkisebb távolságot keressük az összevonáshoz. Ezt a módszert is gyakran használják a klaszterelemzés során.



- A *szekvenciális küszöbérték módszer*nél először kiválasztjuk a klaszter középpontját, majd minden egység, ami a középponttól egy előre meghatározott küszöbértéken belülre esik egy klaszterbe kerül. Ezután új középpontot választunk és csoportosítjuk a fennmaradó egységeket (egy egységet csak egy klaszter-középponttal lehet csoportosítani).
- A *párhuzamos küszöbérték módszer* alkalmazásánál a klaszterek középpontját egyidejűleg választjuk ki, a küszöbértéken belüli egységeket pedig a legközelebb eső középponthez rendeljük.
- Az *optimális felosztás módszerének* az a lényege, hogy a megfigyelési egységeket a folyamat során újra hozzárendeljük más klaszterekhez is, hogy egy általános kritériumot optimalizálhassunk (pl.: adott számú klaszterre a klaszteren belüli távolságok átlagát).

A *lebontó* eljárásban kezdetben egyetlen klaszter létezik, amelybe minden adatpont beletartozik, majd ezt tovább osztjuk. Az újabb klaszterek az előző finomításai lesznek. Az eljárások akkor állnak meg, amikor vagy elérnek egy előre megállapított klaszterszámot, vagy a klaszterek közötti távolság egy előre megállapított mértéknél kisebbé válik.

A *nem hierarchikus klaszterelemzés*nél a belső és külső variancia hányadosát ábrázoljuk a klaszterek számának függvényében és az a pont ahol éles törés látható, a megfelelő klaszterszámra utal. A klaszterek egymáshoz viszonyított méretét is érdemes figyelembe venni. A nagy elemszámú klaszterek mellett az 1-2 elemű klaszternek nincs túl sok értelme.

### ***K-középpontú klaszteranalízis***

A K-középpontú klaszteranalízissel is a vizsgált elemek klaszterbe történő besorolását végezhetjük el. A módszer főbb lépései a következők:

- először adjuk meg a kezdeti klaszterek számát
- minden elemet soroljuk be a hozzá legközelebb eső klaszterbe
- határozzuk meg a klaszterek központját (centroidokat)
- az elemeket úgy soroljuk át, hogy csökkenjen az elemek és a centroidok közötti távolság
- az utolsó két lépést addig ismételjük, amíg a klaszterek nem állandósulnak.

Ez az euklideszi távolságokkal dolgozó algoritmus nagyszámú adat kezelésére képes, azonban csak a kvantitatív (intervallum, vagy arány-skálájú) adatok csoportosítására alkalmas.

A fentiekből látható, hogy a klaszterek kialakításának módszere mellett nagyon fontos az elemezni kívánt objektumok közötti távolság vagy hasonlóság meghatározása. Különböző távolságmérési módszereket alkalmaznak. Ezek közül mutatunk be néhányat.

Az **euklideszi távolság** az egyik legismertebb és legelterjedtebb távolságmérő eljárás. Alapelve a Pitagorasz-tételre alapul. Jelölje  $(x_i, y_i)$  az adatpontokat, és  $d(x,y)$  az adatpontok közötti távolságot, akkor

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

A **négyzetes euklideszi távolság** az euklideszi távolságon alapul, annak a négyzete ( $d^2$ ).

A **Pearson-távolság** is az euklideszi távolságból indul ki, amely az egyes objektumok tulajdonságainak különbségeit az adott tulajdonságok szórásához viszonyítja.

$$d = \sqrt{\sum_{i=1}^n \frac{(x_i - y_i)^2}{s_i^2}}$$

A **négyzetes Pearson-távolság** a Pearson-távolság négyzete ( $d^2$ ).

Az **abszolút eltérés** (Manhattan távolság) hasonlít az euklideszi távolságra, ám a négyzet, illetve a négyzetgyök helyett itt a távolságok abszolút értékét vesszük. Ez a két objektum négyzetrács mentén megtehető útjának hossza:

$$d = \sum_{i=1}^n |x_i - y_i|$$

Ha van rá lehetőség, egy adott probléma megoldásakor sokszor érdemes különböző klaszterező módszereket is alkalmazni, a módszerek ellenőrzésére. Amennyiben kb. ugyanolyan eredményre jutunk mindegyikben, akkor tanácsos elfogadni az eredményeket.

A klaszteranalízis sokszor együtt alkalmazható más matematikai-statisztikai módszerekkel, pl. a faktoranalízis, vagy a főkomponens analízis megelőzheti alkalmazását egy adathalmaz vizsgálatakor.

A feladatok megoldásánál elemzői döntést igényel a változók kiválasztása, a távolságok mérése és a klaszterek összevonása. A megoldás végeredménye lehet numerikus, grafikus, de a csoportosítás jóságát nem tudjuk mérni. Fontos, hogy a hasonló csak a hasonlóval alkothat egy csoportot.

## 15.2. Klaszteranalízis az SPSS-ben

Alkossunk csoportokat a **World95.sav** file-ban található országokból, a hozzájuk rendelt változók értékei alapján. A változók kiválasztásánál ugyanazok a szabályok érvényesek a klaszteranalízisre, mint a faktoranalízisre!

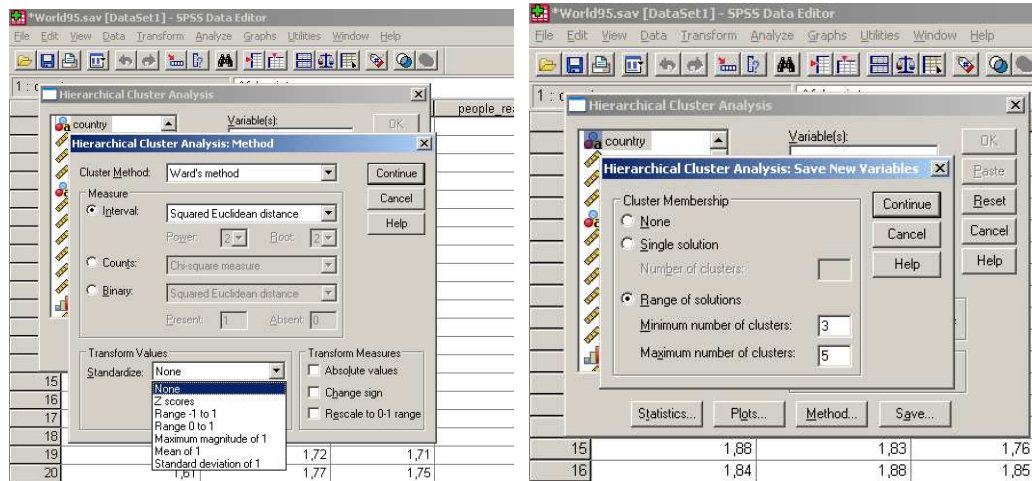
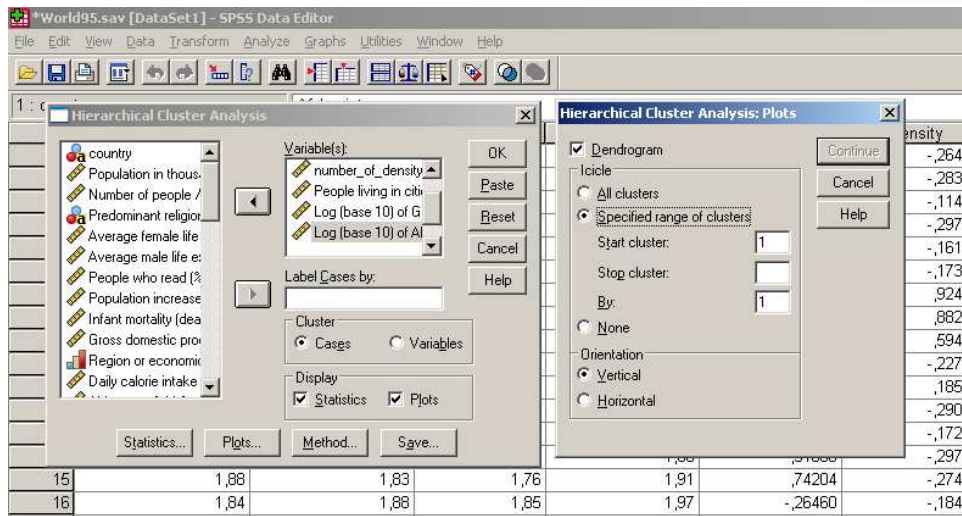
Fontos premissza, hogy a változók száma legyen kisebb, mint az esetek száma. Ha ez nem így van, akkor a változókat faktoranalízissel, főkomponensanalízissel redukálni kell. Módszertanilag egyébként is helyesebb, ha faktoranalízis után hajtjuk végre a klaszteranalízist, ebben az esetben ugyanis változóként használhatjuk a faktorok értékeit, amennyiben elmentettük őket, vagy a faktorokkal legnagyobb korrelációt mutató változókat. Ebben az esetben tehát kevesebb változót használunk fel a csoportképzéshez, és ezek inkább függetleneknek tekinthetők.

Ugyanakkor indokolható az összes változó felhasználása is, abból kiindulva, hogy azzal az adattal kell dolgoznunk, amit nagy nehezen megszereztünk (gondoljunk arra, hogy pénzt adtunk érte), tekintet nélkül arra, hogy fajlagos és abszolút értékek egyaránt lehetnek közöttük, s hogy egyes változók között erős lehet a korreláció, akkor is, ha két külön minőségre/valóságtartamra utalnak (lásd multikollinearitás problémája). Ez az eljárás, bár matematikai szempontból nem a leghelyesebb, gyakorlati megfontolásokból védhető. Nem szerencsés azonban, ha  $a$  és  $b$  változó mellett az  $c$  változó is szerepel, amelyből  $a$  és  $b$  is egyszerűen számítható. Egy faktoranalízis ezeket vélhetően egy faktorba sorolná. Klaszteranalízisben való használatuk tehát azt jelenti, hogy ugyanazt a minőséget/független változót súlyozottan vettük figyelembe (ami nem tilos, csak tisztában kell lenni azzal, hogy más eredményt ad).

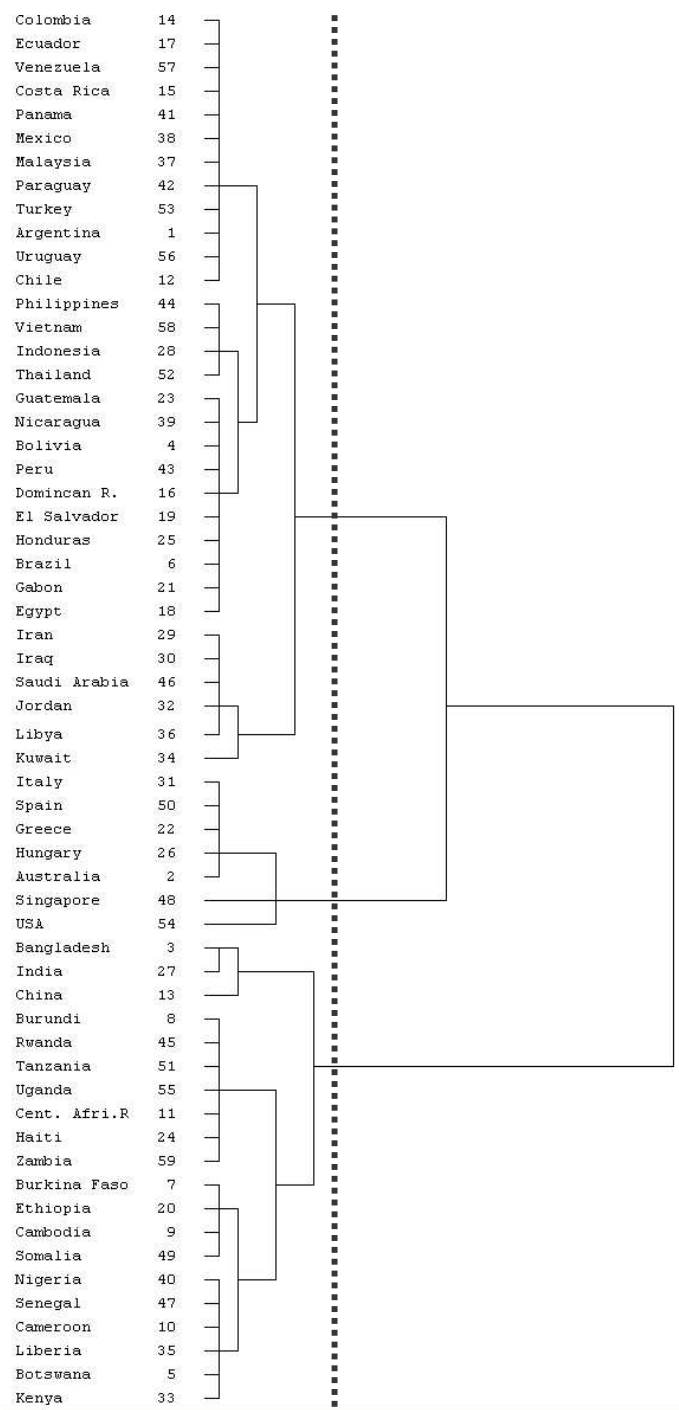
Nem árt, ha a változók normáloszlást mutatnak. A standardizálás is szükséges lehet. A standardizálás viszont nem jelenti automatikusan azt, hogy adataink normáloszlásúvá válnak. Ha nem standardizáljuk adatainkat, az kihat a végeredményre: más egy standardizált és standardizálás nélküli dendrogram megváltozhat az esetek csoportbesorolása.

Az **Analyze – Classify – Hierarchical Cluster** parancsot kiválasztva a párbeszédablakban kiválogathatjuk változóinkat. Ha azt szeretnénk, hogy az egyes esetek könnyen azonosíthatók legyenek, akkor a **Label Cases by** ablakba érdemes az azonosítómezőt betenni (**countries** a **world95.sav** file-ban). Kiválaszthatjuk, hogy az eseteket, vagy a változókat akarjuk klaszterizálni. Az utóbbi eset a faktoranalízisre hasonlít, csak vizuálisabb, de így is elkülöníthetünk független változócsoportokat. Amennyiben a változók klaszterizálását végezzük el, utána egy klaszterből elegendő egy változó beemelése az esetek csoportosítását célzó vizsgálatunkba, ezzel elértük, hogy ugyanaz a „jelenség” csak egyszer szerepeljen.

A **Plots** ablakban kiválaszthatjuk, hogy dendrogramot, vagy Icicle-t akarunk használni, megadhatjuk, hogy az összes klasztert 1-től n-ig akarjuk ábrázolni, vagy csak egy bizonyos intervallumot.



A **Method** ablakban érdemes a **Ward**-féle módszert választani, adatainkat pedig itt lehet standardizálni. Ennek is több módja lehetséges. Itt kell kiválasztani azt is, hogy a sorokon vagy változókon belül történjen a **standardizálás**, ez ugyanis nem mindegy. Ha eseteket klaszterizálunk, akkor szinte mindig a változókat kell standardizálni (kivéve az idősoranalízis: ha a változó az év, akkor az eseteken belül kell standardizálni, így a hasonló trendet mutató országok kerülnek egy csoportba, akkor is, ha az eredeti értékek között nagyságrendi különbség volt).



Az összes változó felhasználásával készített dendrogram és az esetek klaszterekbe sorolása (a pontozott függőleges vonal berajzolásával)

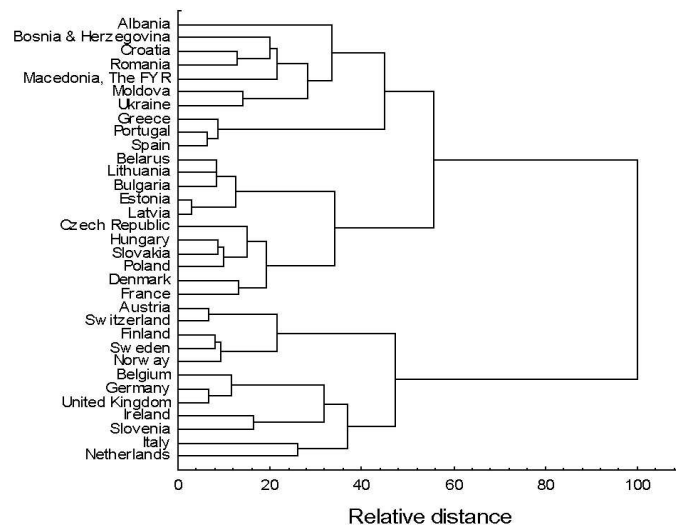


A **Save** ablak lehetővé teszi, hogy az esetek klaszterbe sorolását elment-  
sük. Megadható egy érték, de intervallum is, ebben az esetben több változót fog  
generálni a program. Világos, hogy ugyanaz az eset (itt ország) nem biztos, hogy  
ugyanabba a klaszterbe kerül 3 csoport esetén és 6 csoport létrehozása esetén.

A **Dendrogram** a leglátványosabb kimenet, a legegyszerűbb pedig a **Cluster  
Membership**. Az utóbbi az egyes esetek klaszterbe sorolását mutatja be, az előb-  
bi ágszerkezete lehetővé teszi, hogy tetszés szerint módosítsuk a csoportok szá-  
mát/határait és a csoportba besorolt esetek számát.

Gyakorlásképpen vessük össze dendrogramunkat a faktorok, valamint az  
5 főkomponens (olvasni tudó nők aránya, GDP/fő, lakosság szám, népességnöve-  
kedés, AIDS esetek száma) felhasználásával készített dendrogramunkkal. A  
faktorokat nem kell standardizálni, az 5 változót viszont igen. Látni fogjuk, hogy  
az 5 faktor alapján készült dendrogram esetében az ágak (klasztercsoportok) job-  
ban elkülönülnek, mint az 5 változó alapján készített esetében. Mindhárom kime-  
net különbözik egymástól, tehát a változók (és a módszer) megválasztása erősen  
befolyásolja a kimenetet (ezért fontos megadni a módszertani leírást, ha ilyen  
vizsgálatokat végzünk).

Második példánk az európai országokat mezőgazdaságuk hasonlósága  
alapján klaszterezi. A Nyugat-Balkán és a Mediterráneum jól elkülönül, miként  
egy csoport a Baltikum és Belorusszia. Magyarország a visegrádi országokkal  
került egy csoportba, közel állva Dániához, az EU-támogatásokat illetően nagy  
rivális Franciaországhoz, de a Baltikumhoz is. A fa másik nagy ágát az alpesi és  
nyugat-európai országok kerültek (Németországgal, az Egyesült Királysággal,  
Belgiummal, Svájjal).



változók	adatforrás	változók	adatforrás
Savanyodással fenyegetett mezgazdasági területek % (1990)	Stockholm Environment Institute at York, Acidification in Developing Countries: Ecosystem Sensitivity and the Critical Loads Approach at the Global scale, 2000 via ciesin.org; adapted by NationMaster database, 2004	Szárazföldi terület	FAOSTAT 2004
A mezőgazdaság részesedése a GDP-ből (%) (1990)	CIA World Factbook, December 2003 via NationMaster database, 2004	National Biodiversity Index (fajgazdagság alapján 0-1 között)	WCMC (1992). Development of a national biodiversity index. World conservation Monitoring Centre, Cambridge, UK. Unpublished.; SCBD. 2001. Global Biodiversity Outlook. Secretariat of the Convention on Biological Diversity. Quebec, Canada; <a href="http://www.biodiv.org/gbo">www.biodiv.org/gbo</a>
Művelés alatti terület (ha) (1990)	FAOSTAT 2004; UNECE 2004 reports	Nem őshonos növényfajok sűrűsége (%)	The Flora Europaeae 1997
Művelés alatti terület (%) (2000)	FAOSTAT 2004	szántó természetes kezelés alatt (%) 2001-2003	Konecny 2003; World Resources Institute via NationMaster database, 2004
Mezőgazdasági termelés Indexe (PIN) (1996-98, 1989-hez képest)	FAOSTAT 2004	Peszticid-használat kg/ha	World Resource Institute, World Resources 2000-2001, Washington, DC: WRI, 2000. via ciesin.org & NationMaster database, 2004; Environ in Sloven 1996 Tab P-13
Mezőgazdasági vízhasználat a teljes %-ában (2000)	AQUASTAT 2004, waterreport.pdf 2001; <a href="http://www.seerecon.org">www.seerecon.org</a> ;	Népsűrűség (1993-1997)	compiled from UN reports, and Earthtrends
Szántó a mezgazdaságilag hasznosított terület %-ában (2000)	FAOSTAT 2004	Védett területek (%) IUCN I-IV (1997)	SCBD. 2001. Global Biodiversity Outlook. Secretariat of the Convention on Biological Diversity. Quebec, Canada; <a href="http://www.biodiv.org/gbo">www.biodiv.org/gbo</a> ; Jacaranda Atlas via NationMaster database, 2004
GDP \$ (per fő)	CIA World Factbook, December 2003 via NationMaster database, 2004	Falusi lakosság (2000) (%)	FAOSTAT 2004
GDP növekedése (%) (1975-2000, per fő)	CIA World Factbook, December 2003 via NationMaster database, 2004	Talajszuszpenzió a vízben ln(mg/l) (1994-1996)	United Nations Environment Programme (UNEP), Global Environmental Monitoring System/Water Quality Monitoring System. <a href="http://www.cciw.ca/gems/">http://www.cciw.ca/gems/</a> , with data for an additional 29 countries from Prescott-Allen, R. The Wellbeing of Nations, Washington, DC: Island. via NationMaster database, 2004
Öntözött terület a teljes hasznosított ter. %-ában	FAOSTAT 2004; AQUASTAT 2004	Műtrágya (2001) kg/ha	FAOSTAT 2004

Az európai országok klaszterezése a mezőgazdaságot leíró változók (lásd fent) alapján (LIIRA et al., 2008 szerint)

Harmadik példánkban a klaszteranalízis egy 8 m mély fúrás rétegbeosztását segítette elő, a pH, szervesanyag, mésztartalom, szín (mivel ez nem szám, itt a sötéttség árnyalatai kaptak számértéket egy Munsell-féle skálát használva) és szemeloszlás (homoktartalom %) alapján először 5, majd 8 klaszter létrehozásával finomítva a rétegtant.

Mélység (cm)	Szín és anyag	pH (H <sub>2</sub> O, KCl)	Szervesanyag (%)	CaCO <sub>3</sub> (%)	CLUSTER 5	CLUSTER 8
0-20	yellow sand with signs of soilification	6,98:6,68	30,9%	14,50%	1	1
20-40	greenish grey silty sand with coal stripe	7,06:6,85	11,98%	11,62%	2	2
40-60	greasy yellowish gray clayey silt	7,16:7,01	3,4%	21,1%	2	2
60-80	black, clayey soil, calcareous	7,43:6,99	12,3%	34%	3	3
80-100	sandy clayey silt with coal, clayey-calc. shale clasts	7,01:6,88	27,7%	47,7%	4	4
100-120	sandy clayey silt with coal, clayey-calc. shale clasts	6,96:6,86	20%	39,5%	4	4
120-140	sandy clayey silt with coal, clayey-calc. shale clasts	6,94:6,82	21,6%	47,3%	4	4
140-160	sandy clayey silt with coal, clayey-calc. shale clasts	7,00:6,77	14,6%	23,2%	4	4
160-180	160-170 green clay, 170-180 sandy clay with organic material, consolidated	7,36:6,94	16,1%	29,2%	2	5
180-200	loamy sand with coal and organic material	7,34:6,91	12,3%	39,7%	2	2
200-220	<b>green dark gray silty clay</b>	7,06:7,14	4,7%	43,7%	2	5
220-240	<b>clay, aquitard</b>	7,19:6,98	14,2%	52,8%	4	6
240-260	same, brown	7,66:7,10	8%	43,9%	4	6
260-280	same, black, organic	7,76:7,17	6,7%	47,2%	2	5
280-300	same, organic green	7,60:7,10	6,1%	50,7%	2	5
300-320	<b>grayish green silty clay</b>	7,18:7,07	10,66%	54,6%	2	5
320-340	grayish green sandy clay	7,83:7,25	11,9%	55,4%	2	5
340-360	grayish green sandy clay	7,55:7,36	2,3%	45%	2	5
360-380	same, from 370 brownish red sandy silt	7,63:7,09	2,6%	27,5%	2	5
380-400	brownish black organic sandy silt	6,86:6,54	12,9%	6,3%	3	7
400-420	upper part: black organic (coaly), bottom: limonitic green silty clay	7,17:6,74	12,87	6,5%	3	3
420-440	<b>blue-gray clay reductive, unvented, aquitard</b>	7,53:7,00	12,5%	18%	4	6
440-460	black coal, organic material sandy clay	6,86:6,66	27,4%	3,8%	3	7
460-480	black coal, organic material sandy clay	6,97:6,76	24,4%	26,5%	3	7
480-500	black coal, organic material sandy clay	6,92:6,76	13,7%	17,2%	3	7
500-520	black coal, organic material sandy clay, humid	7,09:6,87	13,89%	25,48%	3	7
520-540	<b>same, but greasy, wet plastic, like plasticine dark gray clay</b>	7,18:6,94	20,11%	25,88%	4	6
540-560	dry, sandy material with claymarl and organic clasts	7,17:6,95	19,50%	36,03%	2	2
560-580	dry, sandy material with claymarl and organic clasts	7,39:7,17	12,03%	34,78%	2	2
580-600	dry, sandy material with claymarl and organic clasts, greenish gray	7,53:7,22	6,20	37,55	3	8
600-620	dry, sandy material with claymarl and organic clasts, greenish gray	7,43:6,88	9,52%	30,93%	2	2
620-640	<b>red, greenish clay, shale, wet, anaerobic</b>	7,31:7,04	9,07%	28,70%	2	5
640-660	calcareous dark gray organic material	7,12:6,86	20,05%	15,64%	4	4
660-680	calcareous dark gray clay with organic material	7,10:6,85	23,37%	31,15%	4	4
680-700	calcareous dark gray clay with organic material	7,08:6,86	17,76%	30,9%	4	4
700-720	<b>greasy, wet, plastic, dark gray clay</b>	7,02:6,86	13,47%	39,58%	4	6
720-740	greenish gray organic material	7,83:6,83	12,55%	36,34%	4	4
740-760	dry, calcareous sandy material with organic content	7,27:6,76	11,73%	50,26%	2	2
760-780	gray sandy-silty material	7,34:7,22	7,13%	61,65%	2	5
780-800	<b>dark gray plastic, dark gray clay</b>	7,15:6,94	11,58%	31,38%	4	6

## Feladatok

- 1, Határozd meg a hasonló rétegeket a fent látható folyamatos mintavételezésű fúrás adatainak felhasználásával.
- 2, Mely meteorológiai állomások jellemezhetők hasonló klímával, az alanti táblázat adatai alapján? Mely változók hasonlítanak egymásra?

Állomás	Csapadék (mm)	Évi átlag-hőmérséklet (°C)	Potenciális párolgás (mm)	Fagymentes napok száma	Napsütéses órák száma
1	3250	25	2530	354	2850
2	1520	23	1560	298	2209
3	325	3	120	50	1650
4	654	14	450	258	1899
5	120	20	400	360	3014
6	985	9	875	185	2005
7	2001	15	1500	299	2006
8	1253	8	863	315	1750

3, Mely szénbányák jellemezhetők hasonló futású termeléssel?

Táró	1995	1996	1997	1998	1999
1	7500	7890	8102	7325	7556
2	2560	2890	3150	2560	2100
3	12098	10403	7850	7563	5699
4	7580	6530	4563	3333	9899
5	12068	12069	15048	16098	16009
6	12058	12048	13058	13050	12009
7	956	1560	2360	1630	1700

## 16. Diszkriminancia-analízis

A diszkriminancia-analízis olyan adatelemzési módszer, amelynél a kvantitatív változók segítségével új változóként csoportokat kívánunk alkotni, illetve a csoportba sorolás helyességét szeretnénk ellenőrizni. A diszkriminancia-analízis alkalmazásával a függő változók értékeit a független változók értékeivel magyarázzuk. Egyrészt az a célunk, hogy a változók közötti összefüggést felfedezzük, másrészt pedig az, hogy a függő változók ismeretlen értékeit a független változók értékei alapján előre megmondjuk.

A diszkriminancia-analízis alkalmazásánál olyan diszkriminancia-függvényt kell meghatározni, amely a „becslő” változók lineáris kombinációjaként a függő változók kategóriáit a legjobban szétválasztja. Közben a változókat figyelembe véve meg kell azt is állapítani, hogy van-e szignifikáns különbség a csoportok között. Fontos azoknak a változóknak a meghatározása, amelyek alapján a csoportok közötti különbség jól magyarázható. A diszkriminancia-analízisben is – a lineáris regresszióhoz hasonlóan – olyan egyenest keresünk, amely a legjobban szétválasztja az elemzendő csoportokat. Természetesen a csoportosítás pontosságának a megállapítása is nélkülözhetetlen.

A változókat néhány ( $H = 1, 2, 3, 4, \text{ stb.}$ ) olyan csoportba osztjuk, amelyek kölcsönösen kizárják egymást. A független változóknak ( $\mathbf{p}$ ) normális eloszlású kvantitatív adatokat kell tartalmazniuk minden csoportban. Minden érték/adat a  $\mathbf{H}$  számú csoport valamelyikébe tartozik és a  $p$  független változóval jellemzett. Az eljárás akkor a leghatékonyabb, ha a csoportképző változó valódi kategoriális változó. Ha a csoportképző változó kvantitatív, akkor az adatokban rejlő információ jobb hasznosítása érdekében célszerűbb lineáris regressziót alkalmazni.

A független (előrejelző) változók alkalmas módon meghatározott lineáris kombinációja az ún. **diszkriminancia-függvény**, amelynek alapján a csoporthoz tartozás megadható a következő képlettel:

$$d = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p$$

Az egyes együtthatók (koefficiensek) becslésére olyan módszer alkalmas, amely egyedi csoportokból indul ki. Figyelembe veszi, ha az egyes csoportok és függvényértékek között kapcsolat van.

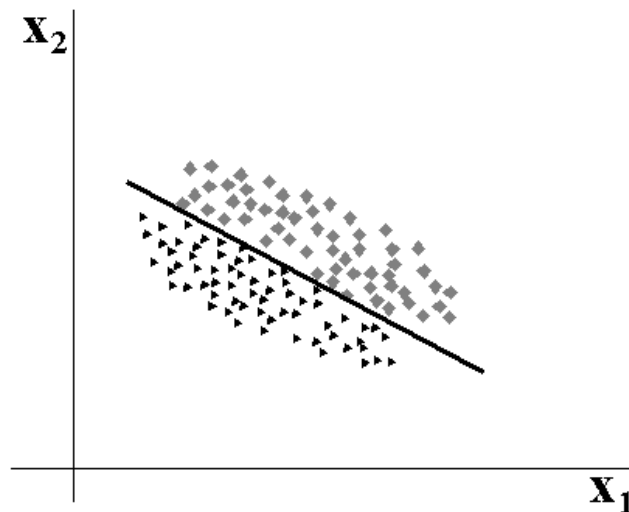
A diszkriminancia-függvénynek a  $\mathbf{b}$  konstansait úgy választjuk meg, hogy a csoportok közötti négyzetösszeget elosztjuk a csoportokon belüli négyzetösszeggel:

$$\frac{\frac{Q_k}{f_k}}{\frac{Q_b}{f_b}} = \frac{Q_k}{f_k} \cdot \frac{f_b}{Q_b} = \frac{Q_k}{H-1} \cdot \frac{n-H}{Q_b}$$

A cél az, hogy ennek az értéke maximális legyen, mert ez adja a lehetséges legjobb diszkriminációt. Ezzel a módszerrel a különböző csoportok függvényértékei a lehető legnagyobb különbségeket adják. Ha a változó adott, akkor ennek a képletnek a használatával azt is vizsgálhatjuk, hogy van-e különbség a csoportok között. Ha a diszkriminancia analízisnél a fenti hányadost választjuk, akkor a vizsgálatot elvégezhetjük F-próbával.

Egy másik használatos statisztikai mutató a 0 és 1 között változó Wilks-féle lambda, amelyet a csoportokon belüli négyzetösszeg és a teljes négyzetösszeg hányadosával határozhatunk meg. Ennek akkor lesz 1 az értéke, ha valamennyi csoport átlaga azonos. Nullához közeli értéket pedig akkor kapunk, ha a csoportokon belüli variabilitás kicsi a teljes variabilitáshoz képest.

Először az egyszerűség kedvéért legyen két csoportunk ( $H = 2$ ). Az ábrán az egyik csoport elemeit háromszöggel, a másikat négyszöggel jelöltük, továbbá  $p = 2$  független változó jele  $x_1$  és  $x_2$ . Megkeressük a két pontthalmazt legjobban elválasztó egyenest. Ez az egyenes már jól diszkriminál, és a diszkriminanciafüggvényt ekkor  $d = b_0 + b_1x_1 + b_2x_2$  alakban írható.



Ha a csoportok száma kettőnél nagyobb, szemléletesen nem mutatható be a diszkriminanciafüggvény származtatása. Ilyen esetekben az ún. *kanonikus változók* segítségével történik az egyes csoportokba tartozó esetek megkülönböztetése.

Az *első kanonikus változó* – az első főkomponenshez hasonlóan – a független változóknak egy olyan lineáris kombinációja, amely egy dimenzióban maximalizálja a  $H$  számú csoport közötti különbséget. A *második kanonikus*

*változó* az elsőtől független lineáris kombináció, amely egy másik dimenzióban maximalizálja a csoportok közötti különbséget, és így tovább...

Láttuk hogy  $H = 2$  esetben egyetlen diszkriminancia (vagy kanonikus) függvény elegendő volt az elemzéshez szükséges kanonikus változók száma általánosan megadva:  $H - 1$  és  $p$  közül a kisebb.

Egy jó diszkriminancia-függvény azzal jellemezhető, hogy az egyes csoportok függvény közéértékei jól elkülönülnek egymástól. Ez a megfontolás képezi az alapját az elemzés jóságának megvizsgálására. A diszkriminancia analízis jóságáról nyerhetünk képet akkor is, ha a diszkriminancia-analízis által feltételezett csoport hovatartozást összehasonlítjuk a valóságos hovatartozással. Ezeket a besorolásokat azután ábrázolhatjuk is, ahol minden területre mindkét függvény értékpárjai jellemzőek, amelyek a mindenkor csoport-hovatartozáshoz vezetnek. A határokat a számok mutatják. A számok egy terület mindenkor csoport-hovatartozását jelentik.

### 16.1. Diszkriminancia-analízis az SPSS-ben

Diszkriminancia-analízist SPSS szoftverben az **Analyse – Classify – Discriminant** menüpontban végezhetünk. Megint a **world95.sav** file felhasználásával mutatjuk be ennek elvi menetét. A file tartalmazza az országokat és a hozzá tartozó változókat: ezek egy része több esetben nem is numerikus változó vagy nominális adat, tehát az országok csoportosítását teszi lehetővé. Ilyen pl. a vallás. Felmerülhet a kérdés, hogy vajon a vallás elég jelentős tényező-e ahhoz, hogy befolyásolja/meghatározza egy ország egyéb jellemvonásait? Másképpen fogalmazva, vajon az országokat jellemző változók értékei alapján következtethetünk-e a vallásra? Ha igen, mely változók lesznek erre a legalkalmasabbak?

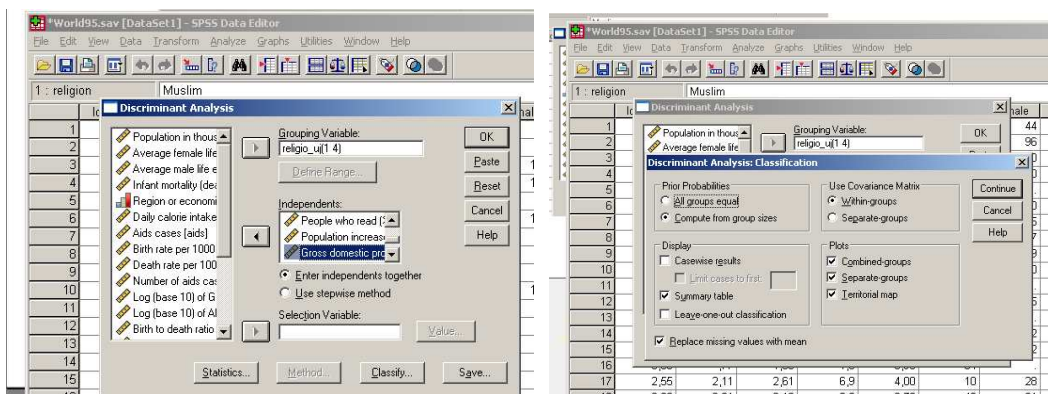
Feladatunkban tehát a változók értékei alapján megkíséreljük az országokat azonosítani a hozzájuk tartozó vallással. Ha a diszkriminancia-analízis eredményeképpen nagy sikerességgel soroljuk be az országokat az eredeti csoportba, akkor megtaláltuk a változókat, ha a visszaosztályozás alacsony hatékonyságot mutat, akkor vagy új változókat kell kipróbálni, vagy feladni a nullhipotézist, hogy egyéb változók értékei alapján, bármely országról megmondható, hogy milyen vallású a lakossága.

Első lépésként újra kell kódolni a valláshoz tartozó értékeket (**Transform – Recode into new variable**), ugyanis a művelethez számokra van szükség (pl. muslim=1, catholic=2, orthodox=3, protstnt=4, stb). Fontos, hogy a neveket pontosan írjuk be!

Ezt követően az **Analyse – Classify – Discriminant** menüpontban a **Grouping variable**-hoz beírjuk a vizsgálandó változót - amelybe az eseteinket visszaosztályozni kívánjuk más változók alapján, tehát most a vallást (**religion**), – majd meghatározzuk a felvehető maximum és minimum értéket (azaz a csoport-

tok számát). A független változókhoz (**independents**) kiválogatunk néhány változót, amelyről azt feltételezzük, hogy alkalmasak csoportok elkülönítésére (pl. a népszaporulat, a GDP, a városi lakosság aránya, stb.). Értelemszerűen kerüljük a szöveges változókat!. Választhatunk az **Enter** és a **Stepwise** módszer között, tehát, hogy a változók egyszerre kerülnek-e a „kalapba”, vagy egymás után. Ha el akarjuk kerülni a **multikollinearitás** jelenségét mutató változók használatát, akkor előtte faktoranalízissel kivont faktorokat, vagy a faktorról legnagyobb korrelációt mutató változót is használhatjuk. Az „együtt mozgó”, hasonló értékeket mutató változók használata ugyanis súlyozást jelent, ami viszont elvben növelheti is a visszaosztályozás sikerességét, mint ahogy csökkentheti is!

A **Classify** gomb megnyomásával további beállítások eszközölhetők: megadható, hogy egyforma-e a csoportok nagysága (**All groups equal, Compute from group size**), lekérhetjük az újraosztályozást összegző táblázatot (**Summary table**), és a diagramokat (**Plots**). A diszkriminancia-analízis során legyártott függvények (2, esetleg 3) értékei alapján az egyes esetek felkerülnek 2D egy pontdiagramra (koordináta-rendszer), a rájuk jellemző csoportcentrummal (**Plots-Combined groups**). A diagramról a többi csoportcentrumtól való távolság is leolvasható, miképp az is, hogy mennyire különülnek el az egyes csoportok, ill. mekkora közöttük az átfedés. Akkor lesz sikeres a visszaosztályozás - tehát akkor választottunk jó változókat -, ha a csoportcentrumok egymástól távol, az esetek viszont a csoportcentrumhoz közel helyezkednek el, és a csoportok nem fedik át egymást. Szintén lekérhető, hogy a csoportok hogyan osztják fel egymás között a 2D fiktív teret (**Territorial map**).

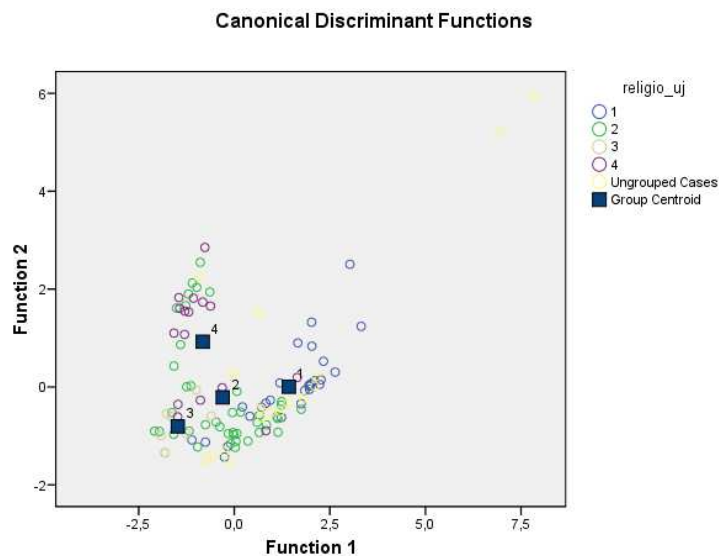


A **Statistics** menü kiválasztásával a korrelációs mátrixot, az ANOVA-analízist, a létrehozandó diszkriminancia-függvény koefficienseit és a diszkriminancia-analízis megbízhatóságát vizsgálhatjuk meg. A **Save** gomb megnyomásával az adatbázisba új változóként elmenthetjük a valószínűsített csoport-



besorolást (**Predicted group membership**), és annak valószínűségét (**Probabilities**).

Az egyszerűség kedvéért csak négy vallás esetében végeztük el a vizsgálatot, s nem használtuk fel az összes változót sem. Példánkban látszik, hogy a választott változók csoportcentruma egymáshoz közel esett, a csoportok esetei összekeverednek, nem különülnek el, ennek köszönhetően, az eredeti 1-es csoportba 27 esetből 20 került helyesen visszasorolásra (74%-os siker), 7 viszont a 2. csoportba (tehát 7 muszlim országot katolikusnak minősítettünk a választott változók alapján). A 2. csoportból 5 ország került az 1-es, 1 a 3. és 8 a 4. csoportba (66%-os sikeresség). Összességében az esetek 65%-a helyesen került visszaosztályozásra, ami meghaladja a véletlen valószínűséget (50%), de más változók kiválasztásával ez az érték talán növelhető.



**Classification Results<sup>a</sup>**

		religio_uj	Predicted Group Membership				Total
			1,00	2,00	3,00	4,00	
Original	Count	1,00	20	7	0	0	27
		2,00	5	27	1	8	41
		3,00	0	5	3	0	8
		4,00	1	5	0	10	16
		Ungrouped cases		9	7	0	1
%		1,00	74,1	25,9	,0	,0	100,0
		2,00	12,2	65,9	2,4	19,5	100,0
		3,00	,0	62,5	37,5	,0	100,0
		4,00	6,3	31,3	,0	62,5	100,0
		Ungrouped cases		52,9	41,2	,0	5,9

a. 65,2% of original grouped cases correctly classified.

### ***Feladatok***

- 1,** A world95.sav állomány adatainak vizsgálatával elemezze, hogy a világ országainak
  - a, vallási hovatartozása
  - b, gazdasági fejlettsége (region or economic group)

megállapítható-e más változók értékei alapján (milyen sikeres a visszaosztályozás a vallási csoportokba/gazdasági régióba más változók felhasználásával)? Mely változók mutatják a legjobb visszaosztályozási arányt (mely változók alapján a legsikeresebb a vallás/gazdasági régió beazonosítása)?

- 2,** Vajon a klíma mennyire tükrözi ez egyes országok fejlettségét? A földrajzi determinizmus elméletének van-e létjogosultsága? (Azonosítható-e a klímatípus fejlettségi mutatók alapján?)

## Felhasznált irodalom

- Geiger J. (2007): Geomatematika. Szegedi Egyetemi Kiadó. JATEPress, 116. p.
- Herendi I. - Lengyel I. - Nemes Nagy J. - Sikos T. T. - Szónokyné Ancsin G. - Szörényiné Kukorelly I. Szörényi M. (1996): Társadalomföldrajzi elemzések számítógépen. JATEPress, Szeged, 194. p.
- Ketskeméty L. - Izsó L. (2005): Bevezetés az SPSS programrendszerbe. Bp. ELTE, Eötvös Kiadó 459.p
- Köves P. - Párniczky G. (1981): Általános statisztika I-II. Közgazdasági és Jogi Könyvkiadó. Bp., 362+387.p.
- Liira, J. - Aavik, T. - Parrest, O. - Zobel, M. Agricultural sector, rural environment and biodiversity in the Central and Eastern European EU member states. In: Acta Geographica Debrecina, Landscape and Environment Series. Vol. 2. 46-65. pp.
- Móricz F. - Abonyi Gy-né (1975): Matematikai módszerek a földrajzban. Tankönyvkiadó, Bp., 285. p. (kézirat)
- Nemes Nagy J. (szerk.) (2005): Regionális elemzési módszerek. ELTE, Regionális Földrajzi Tanszék - MTA Regionális Tudományi Kutatócsoport. 284. p.
- Péczely Gy. (1979): Éghajlat. Tankönyvkiadó, Bp., 336. p.
- Pénzes J. (2007): A területi jövedelemegyenlőtlenségek alakulása az észak-alföldi régióban. In: Területi Statisztika, 4. 358-379. pp.
- Sajtos L. – Mitev, A. (2007): SPSS kutatási és adatelemzési kézikönyv. Bp. Alinea. 402.p.
- Süli-Z. I. (1980): A Hegyköz faluföldrajzi vizsgálata. Debrecen, Studia Geographica 3.
- Szénás Gy.: Geofizikai teleptan. Akadémiai Kiadó, 1958.
- Szunyogh G. (1999): Változó valóság. Fejezetek a dinamikus földrajz tárgyköréből. Oskar Kiadó, Szombathely, 201.p.
- [http://www.agt.bme.hu/staff\\_h/varga/vetulet.html](http://www.agt.bme.hu/staff_h/varga/vetulet.html)
- [http://psycho.unideb.hu/munkatarsak/math\\_janos/statisztika\\_II/nempar.doc](http://psycho.unideb.hu/munkatarsak/math_janos/statisztika_II/nempar.doc)
- [http://hu.wikipedia.org/wiki/K%C3%A9rtmint%C3%A1s\\_t-pr%C3%B3ba](http://hu.wikipedia.org/wiki/K%C3%A9rtmint%C3%A1s_t-pr%C3%B3ba)
- <http://www.stud.u-szeged.hu/Batori.Csaba/AlkStat.pdf>
- <http://www.google.hu/search?q=korrel%C3%A1ci%C3%B3sz%C3%A1m%C3%ADt%C3%A1s&hl=hu&start=10&sa=N>
- [http://psycho.unideb.hu/statisztika/pages/p\\_4\\_1.xml](http://psycho.unideb.hu/statisztika/pages/p_4_1.xml)



# MELLÉKLETEK

## 1. Hosszmértékek

Mértékegység	Méret (m)
<i>Ószövetségi hosszértékek</i>	
1 mérónád = 6 könyök	kb. 270 cm
1 könyök	kb. 45 cm
1 régi könyök	kb. 52,5 cm
1 arasz = 3 tenyér	kb. 22,5 cm
1 tenyér = 4 ujj	kb. 75 mm
1 ujj	kb. 18,75 mm
<i>Újszövetségi hosszértékek</i>	
1 mérföld (ezer lépés)	1478 m
1 stádium	185 m
1 öl	1,85 m
1 könyök	55 cm
1 láb	30 cm
<i>Ógörög hosszértékek</i>	
1 szkhoinosz = 60 ógörög sztadion	11116,8 m
1 ógörög sztadion = 6 plethron	185,28 m
1 plethron = 100 pusz (láb)	30,88 m
1 pusz (láb) = 16 daktülosz (hüvelyk)	0,3088 m
1 daktülosz (hüvelyk)	0,0193 m
1 orgüa (öl) = 6 pusz (láb) = 4 pekhüsz (könyök) = 24 paleiszte (tenyér) = 96 daktülosz (hüvelyk)	1,8528 m
1 pekhüsz (könyök)	0,4832 m
1 paleiszte (tenyér)	0,0772 m
1 attikai sztadion	164 m
1 olimpiai sztadion	190–192 m
<i>Osztrák hosszértékek</i>	
bécsi hüvelyk	0,0263 m

bécsi láb = 12 bécsi hüvelyk	0,3160 m
bécsi róf = 29 hüvelyk	0,7775 m
bécsi öl = 6 bécsi láb	1,8964 m
bécsi mérföld = 4000 bécsi öl	7585,92 m
osztrák posta mérföld = 4000 öl	7585,94 m
<b><i>Magyar hosszértékek</i></b>	
budai öl	2,9790 m
pozsonyi öl	1,9010 m
selmeci öl	2,0220 m
bányaöl	2,0258 m
magyar mérföld	8353,6 m
magyar földrajzi mérföld	9276 m
magyar kis földrajzi mérföld	1855 m
<b><i>Angol hosszértékek</i></b>	
1 league = 3 statute mile (szárazföldi)	4,828 km
1 league = 3 nautical mile (tengeri)	5,565 km
1 nautical mile = 2026 yard	1853,184 m
1 statute mile = 8 furlong = 1760 yard	1609,344 m
1 furlong = 40 pole	201,16 m
1 pole/perch/rod = 5,5 yard	5,03 m
1 inch = 10 vonás (hüvelyk)	2,54 cm
1 line (vonás)	2,54 mm
1 mile = 1760 yard	1,609 km
1 fathom = 2 yard	1,83 m
1 yard = 3 foot	91,44 cm
1 foot = 12 inch	30,48 cm
1 inch = 10 line	2,54 cm

<i>Egyéb hossz mértékek</i>	
1 tengeri mérföld	1852 m
fonál (tengerészeti) = 6 foot	1,829 m
francia mérföld	4452 m
német földrajzi mérföld	7420 m
1 hüvelyk = 2,0-3,1 <u>cm</u> németül Zoll	
angol hüvelyk	2,54 cm
bánya hüvelyk, bécsi hüvelyk	2,63 cm
erdélyi hüvelyk, erődítmény hüvelyk , királyi hüvelyk, mérnöki hüvelyk , pozsonyi hüvelyk	2,60 cm

## 2. Területmértékek

Mértékegység	Méret
1 iga (egy iga ökörrrel naponta felszántható terület)	kb. 0,25 ha = 2500 m <sup>2</sup>
1 katasztrális hold = 1600 négyszögöl	0,5760 ha
1 magyar hold = 1200 négyszögöl	0,4320 ha
1 négyszögöl	3,6 m <sup>2</sup>
bécsi négyszögöl = 36 négyszögláb	3,696 m <sup>2</sup> = 0,03596 ár
1 arura	0,276 hektár = 276 m <sup>2</sup>
1 plethron	0,095 hektár = 95 m <sup>2</sup>
1 square mile = 640 acre (négyzetmérföld)	2,590 km <sup>2</sup>
1 acre = 4840 square yard	4046,9 m <sup>2</sup>
1 square = 100 square foot (négyzetláb)	9,290 m <sup>2</sup>
1 square yard = 9 square foot	0,836 m <sup>2</sup>
1 square foot = 144 square inch	0,093 m <sup>2</sup> = 929,01 cm <sup>2</sup>
1 square inch (hüvelyk)	6,45 cm <sup>2</sup>

*A korrelációs együttható valószínűségi szintjei*

<i>f</i>	<i>Valószínűségek</i>				
	<i>0,1</i>	<i>0,05</i>	<i>0,02</i>	<i>0,01</i>	<i>0,001</i>
1	0,98769	0,99692	0,999507	0,999877	0,9999988
2	,90000	,95000	,98000	,990000	,99900
3	,8054	,8783	,93433	,95873	,99116
4	,7293	,8114	,8822	,91720	,97406
5	,6694	,7545	,8329	,8745	,95074
6	,6215	,7067	,7887	,8343	,92493
7	,5822	,6664	,7498	,7977	,8982
8	,5494	,6319	,7155	,7646	,8721
9	,5214	,6021	,6851	,7348	,8471
10	,4973	,5760	,6581	,7079	,8233
11	,4762	,5529	,6339	,6835	,8010
12	,4575	,5324	,6120	,6614	,7800
13	,4409	,5139	,5923	,6411	,7603
14	,4259	,4973	,5742	,6226	,7420
15	,4124	,4821	,5577	,6055	,7246
16	,4000	,4683	,5425	,5897	,7084
17	,3887	,4555	,5285	,5751	,6932
18	,3783	,4438	,5155	,5614	,6787
19	,3687	,4329	,5034	,5487	,6652
20	,3598	,4227	,4921	,5368	,6524
25	,3233	,3809	,4451	,4869	,5974
30	,2960	,3494	,4093	,4487	,5541
35	,2746	,3246	,3810	,4182	,5189
40	,2573	,3044	,3578	,3932	,4896
45	,2428	,2875	,3384	,3721	,4648
50	,2306	,2732	,3218	,3541	,4433
60	,2108	,2500	,2948	,3248	,4078
70	,1954	,2319	,2737	,3017	,3799
80	,1829	,2172	,2565	,2830	,3568
90	,1726	,2050	,2422	,2673	,3375
100	,1638	,1946	,2301	,2540	,3211



### A t-eloszlás táblázata

<i>f</i>	Valószínűségek												
	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,001
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,924
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	1,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	0,126	0,254	0,387	0,527	0,679	0,848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
:	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

Az *f*-eloszlás táblázata  $p = 0,5\%$

$f_2$		A számláló szabadságfoka																			
		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	:	
A számláló szabadságfoka	1	16200	20000	21600	22500	23100	23400	23700	23900	24100	24200	24400	24600	24800	24900	25000	25100	25300	25400	25500	
	2	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	200
	3	55,6	49,8	47,5	46,2	45,4	44,8	44,4	44,1	43,9	43,7	43,4	43,1	42,8	42,6	42,5	42,3	42,1	42,0	41,8	
	4	31,3	26,3	24,3	23,2	22,5	22,0	21,6	21,4	21,1	21,0	20,7	20,4	20,2	20,0	19,9	19,8	19,6	19,5	19,3	
	5	22,8	18,3	16,5	15,6	14,9	14,5	14,2	14,0	13,8	13,6	13,4	13,1	12,9	12,8	12,7	12,5	12,4	12,3	12,1	
	6	18,6	14,5	12,9	12,0	11,6	11,5	10,8	10,6	10,4	10,1	10,0	9,81	9,59	9,47	9,36	9,24	9,12	9,00	8,88	
	7	16,2	12,4	10,9	10,1	9,52	9,16	8,89	8,68	8,51	8,38	8,18	7,97	7,75	7,65	7,53	7,42	7,31	7,19	7,08	
	8	14,7	11,0	9,60	8,81	8,30	7,95	7,69	7,50	7,34	7,21	7,01	6,81	6,61	6,50	6,40	6,29	6,18	6,06	5,95	
	9	13,6	10,1	8,72	7,96	7,47	7,13	6,88	6,69	6,54	6,42	6,23	6,03	5,83	5,73	5,62	5,52	5,41	5,30	5,19	
	10	12,8	9,43	8,08	7,34	6,87	6,54	6,30	6,12	5,97	5,85	5,66	5,47	5,27	5,17	5,07	4,97	4,86	4,75	4,64	
	11	12,2	8,91	7,60	6,88	6,42	6,10	5,86	5,68	5,54	5,42	5,24	5,05	4,86	4,76	4,65	4,55	4,44	4,34	4,23	
	12	11,8	8,51	7,23	6,52	6,07	5,76	5,52	5,35	5,20	5,09	4,91	4,72	4,53	4,43	4,33	4,23	4,12	4,01	3,90	
	13	11,4	8,19	6,93	6,23	5,79	5,48	5,25	5,08	4,94	4,82	4,64	4,46	4,27	4,17	4,07	3,97	3,87	3,76	3,65	
	14	11,1	7,92	6,68	6,00	5,56	5,26	5,03	4,86	4,72	4,60	4,43	4,25	4,06	3,96	3,86	3,76	3,66	3,55	3,44	
	15	10,8	7,70	6,48	5,80	5,37	5,07	4,85	4,67	4,54	4,42	4,25	4,07	3,88	3,79	3,69	3,58	3,48	3,37	3,26	
	16	10,6	7,51	6,30	5,64	5,21	4,91	4,69	4,52	4,38	4,27	4,10	3,92	3,73	3,64	3,54	3,44	3,33	3,22	3,11	
	17	10,4	7,35	6,16	5,50	5,07	4,78	4,56	4,39	4,25	4,14	3,97	3,79	3,61	3,51	3,41	3,31	3,21	3,10	2,98	
	18	10,2	7,21	6,03	5,37	4,96	4,66	4,44	4,28	4,14	4,03	3,86	3,68	3,50	3,40	3,30	3,20	3,10	2,99	2,87	
	19	10,1	7,09	5,92	5,27	4,85	4,56	4,34	4,18	4,04	3,93	3,76	3,59	3,40	3,31	3,21	3,11	3,00	2,89	2,78	
	20	9,94	6,99	5,82	5,17	4,76	4,47	4,26	4,09	3,96	3,85	3,68	3,50	3,32	3,22	3,12	3,02	2,92	2,81	2,69	
	21	9,83	6,89	5,73	5,09	4,68	4,39	4,18	4,01	3,88	3,77	3,60	3,43	3,24	3,15	3,05	2,95	2,84	2,73	2,61	
	22	9,73	6,81	5,65	5,02	4,61	4,32	4,11	3,94	3,81	3,70	3,54	3,36	3,18	3,08	2,98	2,88	2,77	2,66	2,55	
	23	9,63	6,73	5,58	4,95	4,54	4,26	4,05	3,88	3,75	3,64	3,47	3,30	3,12	3,02	2,92	2,82	2,71	2,60	2,48	
	24	9,55	6,66	5,52	4,89	4,49	4,20	3,99	3,83	3,69	3,59	3,42	3,25	3,06	2,97	2,87	2,77	2,66	2,55	2,43	
	25	9,48	6,60	5,46	4,84	4,43	4,15	3,94	3,78	3,64	3,54	3,37	3,20	3,01	2,92	2,82	2,72	2,61	2,50	2,38	
	26	9,41	6,54	5,41	4,79	4,38	4,10	3,89	3,73	3,60	3,49	3,33	3,15	2,97	2,87	2,77	2,67	2,56	2,45	2,33	
	27	9,34	6,49	5,36	4,74	4,34	4,06	3,85	3,69	3,56	3,45	3,28	3,11	2,93	2,83	2,73	2,63	2,52	2,41	2,29	
	28	9,28	6,44	5,32	4,70	4,30	4,02	3,81	3,65	3,52	3,41	3,25	3,07	2,89	2,79	2,69	2,59	2,48	2,37	2,25	
29	9,23	6,40	5,28	4,66	4,26	3,98	3,77	3,61	3,48	3,38	3,21	3,04	2,86	2,76	2,66	2,56	2,45	2,33	2,21		
30	9,18	6,35	5,24	4,62	4,23	3,95	3,74	3,58	3,45	3,34	3,18	3,01	2,82	2,73	2,63	2,52	2,42	2,30	2,18		
40	8,83	6,07	4,98	4,37	3,99	3,71	3,51	3,35	3,22	3,12	2,95	2,78	2,60	2,50	2,40	2,30	2,18	2,06	1,93		
60	8,49	5,79	4,73	4,14	3,76	3,49	3,29	3,13	3,01	2,90	2,74	2,57	2,39	2,29	2,19	2,08	1,96	1,83	1,69		
120	8,18	5,54	4,50	3,92	3,55	3,28	3,09	2,93	2,81	2,71	2,54	2,37	2,19	2,09	1,98	1,87	1,75	1,61	1,43		
:	7,88	5,30	4,28	3,72	3,35	3,09	2,90	2,74	2,62	2,52	2,36	2,19	2,00	1,90	1,79	1,67	1,53	1,36	1,00		

Az  $f$ -eloszlás táblázata  $p = 1\%$

$f_2$	$f_1$	A számláló szabadságfoka																		
		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	:
A számláló szabadságfoka	1	4050	5000	5400	5620	5760	5860	5930	5980	6020	6060	6110	6160	6210	6230	6260	6290	6310	6340	6370
	2	98,5	99,0	99,2	99,2	99,3	99,3	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,5	99,5	99,5	99,5	99,5	99,5
	3	34,1	30,8	29,5	28,7	28,2	27,9	27,7	27,5	27,3	27,2	27,1	26,9	26,7	26,6	26,5	26,4	26,3	26,2	26,1
	4	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,5	14,4	14,2	14,0	13,9	13,8	13,7	13,7	13,6	13,5
	5	16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,3	10,2	10,1	9,89	9,72	9,55	9,47	9,38	9,29	9,20	9,11	9,02
	6	13,7	10,9	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,72	7,56	7,40	7,31	7,23	7,14	7,06	6,97	6,88
	7	12,2	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,47	6,31	6,16	6,07	5,99	5,91	5,82	5,74	5,65
	8	11,3	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,67	5,52	5,36	5,28	5,20	5,12	5,03	4,95	4,86
	9	10,6	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,11	4,96	4,81	4,73	4,65	4,57	4,48	4,40	4,31
	10	10,0	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,71	4,56	4,41	4,33	4,25	4,17	4,08	4,00	3,91
	11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,40	4,25	4,10	4,02	3,94	3,86	3,78	3,69	3,60
	12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,16	4,01	3,86	3,78	3,70	3,62	3,54	3,45	3,36
	13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	3,96	3,82	3,66	3,59	3,51	3,43	3,34	3,25	3,17
	14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,80	3,66	3,51	3,43	3,35	3,27	3,18	3,09	3,00
	15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,67	3,52	3,37	3,29	3,21	3,13	3,05	2,96	2,87
	16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,55	3,41	3,26	3,18	3,10	3,02	2,93	2,84	2,75
	17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,46	3,31	3,16	3,08	3,00	2,92	2,83	2,75	2,65
	18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,37	3,23	3,08	3,00	2,92	2,84	2,75	2,66	2,57
	19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,30	3,15	3,00	2,92	2,84	2,76	2,67	2,58	2,49
	20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,23	3,09	2,94	2,86	2,78	2,69	2,61	2,52	2,42
	21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,17	3,03	2,88	2,80	2,72	2,64	2,55	2,46	2,36
	22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,12	2,98	2,83	2,75	2,67	2,58	2,50	2,40	2,31
	23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,07	2,93	2,78	2,70	2,62	2,54	2,45	2,35	2,26
	24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,03	2,89	2,74	2,66	2,58	2,49	2,40	2,31	2,21
	25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13	2,99	2,85	2,70	2,62	2,54	2,45	2,36	2,27	2,17
	26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	2,96	2,81	2,66	2,58	2,50	2,42	2,33	2,23	2,13
	27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06	2,93	2,78	2,63	2,55	2,47	2,38	2,29	2,20	2,10
	28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,90	2,75	2,60	2,52	2,44	2,35	2,26	2,17	2,06
	29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00	2,87	2,73	2,57	2,49	2,41	2,33	2,23	2,14	2,03
	30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,84	2,70	2,55	2,47	2,39	2,30	2,21	2,11	2,01
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,66	2,52	2,37	2,29	2,20	2,11	2,02	1,92	1,80	
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,50	2,35	2,20	2,12	2,03	1,94	1,84	1,73	1,60	
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47	2,34	2,19	2,03	1,95	1,86	1,76	1,66	1,53	1,38	
:	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,18	2,04	1,88	1,79	1,70	1,59	1,47	1,32	1,00	

Az  $f$ -eloszlás táblázata  $p = 2,5\%$

$f_2$	$f_1$	A számláló szabadságfoka																		
		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	:
A számláló szabadságfoka	1	648	800	864	900	922	937	948	957	963	969	977	985	993	997	1000	1010	1010	1010	1020
	2	38,5	39,0	39,2	39,2	39,3	39,3	39,4	39,4	39,4	39,4	39,4	39,4	39,4	39,5	39,5	39,5	39,5	39,5	39,5
	3	17,4	16,0	15,4	15,1	14,9	14,7	14,6	14,5	14,5	14,4	14,3	14,3	14,2	14,1	14,1	14,0	14,0	13,9	13,9
	4	12,2	10,6	9,98	9,60	9,36	9,20	9,07	8,98	8,90	8,84	8,75	8,66	8,56	8,51	8,46	8,41	8,36	8,31	8,26
	5	10,0	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,68	6,62	6,52	6,43	6,33	6,28	6,23	6,18	6,12	6,07	6,02
	6	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60	5,52	5,46	5,37	5,27	5,17	5,12	5,07	5,01	4,96	4,90	4,85
	7	8,07	6,54	5,89	5,52	5,29	5,12	4,99	4,90	4,82	4,76	4,67	4,57	4,47	4,42	4,36	4,31	4,25	4,20	4,14
	8	7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43	4,36	4,30	4,20	4,10	4,00	3,95	3,89	3,84	3,78	3,73	3,67
	9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	4,03	3,96	3,87	3,77	3,67	3,61	3,56	3,51	3,45	3,39	3,33
	10	6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,85	3,78	3,72	3,62	3,52	3,42	3,37	3,31	3,26	3,20	3,14	3,08
	11	6,72	5,26	4,63	4,28	4,04	3,88	3,76	3,66	3,59	3,53	3,43	3,33	3,23	3,17	3,12	3,06	3,00	2,94	2,88
	12	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,44	3,37	3,28	3,18	3,07	3,02	2,96	2,91	2,85	2,79	2,72
	13	6,41	4,97	4,35	4,00	3,77	3,60	3,48	3,39	3,31	3,25	3,15	3,05	2,95	2,89	2,84	2,78	2,72	2,66	2,60
	14	6,30	4,86	4,24	3,89	3,66	3,50	3,38	3,29	3,21	3,15	3,05	2,95	2,84	2,79	2,73	2,67	2,61	2,55	2,49
	15	6,20	4,77	4,15	3,80	3,58	3,41	3,29	3,20	3,12	3,06	2,96	2,86	2,76	2,70	2,64	2,59	2,52	2,46	2,40
	16	6,12	4,69	4,08	3,73	3,50	3,34	3,22	3,12	3,05	2,99	2,89	2,79	2,68	2,63	2,57	2,51	2,45	2,38	2,32
	17	6,04	4,62	4,01	3,66	3,44	3,28	3,16	3,06	2,98	2,92	2,82	2,72	2,62	2,56	2,50	2,44	2,38	2,32	2,25
	18	5,98	4,56	3,95	3,61	3,38	3,22	3,10	3,01	2,93	2,87	2,77	2,67	2,56	2,50	2,44	2,38	2,32	2,26	2,19
	19	5,92	4,51	3,90	3,56	3,33	3,17	3,05	2,96	2,88	2,82	2,72	2,62	2,51	2,45	2,39	2,33	2,27	2,20	2,13
	20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,84	2,77	2,68	2,57	2,46	2,41	2,35	2,29	2,22	2,16	2,09
	21	5,83	4,42	3,82	3,48	3,25	3,09	2,97	2,87	2,80	2,73	2,64	2,53	2,42	2,37	2,31	2,25	2,18	2,11	2,04
	22	5,79	4,38	3,78	3,44	3,22	3,05	2,93	2,84	2,76	2,70	2,60	2,50	2,39	2,33	2,27	2,21	2,14	2,08	2,00
	23	5,75	4,35	3,75	3,41	3,18	3,02	2,90	2,81	2,73	2,67	2,57	2,47	2,36	2,30	2,24	2,18	2,11	2,04	1,97
	24	5,72	4,32	3,72	3,38	3,15	2,99	2,87	2,78	2,70	2,64	2,54	2,44	2,33	2,27	2,21	2,15	2,08	2,01	1,94
	25	5,69	4,29	3,69	3,35	3,13	2,97	2,85	2,75	2,68	2,61	2,51	2,41	2,30	2,24	2,18	2,12	2,05	1,98	1,91
	26	5,66	4,27	3,67	3,33	3,10	2,94	2,82	2,73	2,65	2,59	2,49	2,39	2,28	2,22	2,16	2,09	2,03	1,95	1,88
	27	5,63	4,24	3,65	3,31	3,08	2,92	2,80	2,71	2,63	2,57	2,47	2,36	2,25	2,19	2,13	2,07	2,00	1,93	1,85
	28	5,61	4,22	3,63	3,29	3,06	2,90	2,78	2,69	2,61	2,55	2,45	2,34	2,23	2,17	2,11	2,05	1,98	1,91	1,83
	29	5,59	4,20	3,61	3,27	3,04	2,88	2,76	2,67	2,59	2,53	2,43	2,32	2,21	2,15	2,09	2,03	1,96	1,89	1,81
	30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,57	2,51	2,41	2,31	2,20	2,14	2,07	2,01	1,94	1,87	1,79
40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,45	2,39	2,29	2,18	2,07	2,01	1,94	1,88	1,80	1,72	1,64	
60	5,29	3,93	3,34	3,01	2,79	2,63	2,51	2,41	2,33	2,27	2,17	2,06	1,94	1,88	1,82	1,74	1,67	1,58	1,48	
120	5,15	3,80	3,23	2,89	2,67	2,52	2,39	2,30	2,22	2,16	2,05	1,94	1,82	1,76	1,69	1,61	1,53	1,43	1,31	
:	5,02	3,69	3,12	2,79	2,57	2,41	2,29	2,19	2,11	2,05	1,94	1,83	1,71	1,64	1,57	1,48	1,39	1,27	1,00	

Az  $f$ -eloszlás táblázata  $p = 5\%$

$f_2$	$f_1$	A számláló szabadságfoka																			
		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	:	
A számláló szabadságfoka	1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254	
	2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5	19,5	19,5
	3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53	
	4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63	
	5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,36	
	6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67	
	7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23	
	8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93	
	9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71	
	10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54	
	11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40	
	12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30	
	13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21	
	14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13	
	15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07	
	16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01	
	17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,49	2,23	2,19	2,15	2,10	2,06	2,01	1,96	
	18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92	
	19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88	
	20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84	
	21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81	
	22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84	1,78	
	23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,13	2,05	2,00	1,96	1,91	1,86	1,81	1,76	
	24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79	1,73	
	25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71	
	26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,07	1,99	1,95	1,90	1,85	1,80	1,75	1,69	
	27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,13	2,06	1,97	1,93	1,88	1,84	1,79	1,73	1,67	
	28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,04	1,96	1,91	1,87	1,82	1,77	1,71	1,65	
	29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,03	1,94	1,90	1,85	1,81	1,75	1,70	1,64	
	30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68	1,62	
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58	1,51		
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47	1,39		
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35	1,25		
:	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22	1,00		

Az *f*-eloszlás táblázata *p* = 10 %

<i>f</i> <sub>2</sub>	<i>f</i> <sub>1</sub>	A számláló szabadságfoka																		
		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	:
A számláló szabadságfoka	1	39,9	49,5	53,6	55,8	57,2	58,2	58,9	59,4	59,9	60,2	60,7	61,2	61,7	62,0	62,3	62,5	62,8	63,1	63,3
	2	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38	9,39	9,41	9,42	9,44	9,45	9,46	9,47	9,47	9,48	9,49
	3	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24	5,23	5,22	5,20	5,18	5,18	5,17	5,16	5,15	5,14	5,13
	4	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94	3,92	3,90	3,87	3,84	3,83	3,82	3,80	3,79	3,78	3,76
	5	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32	3,30	3,27	3,24	3,21	3,19	3,17	3,16	3,14	3,12	3,10
	6	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96	2,94	2,90	2,87	2,84	2,82	2,80	2,78	2,76	2,74	2,72
	7	3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72	2,70	2,67	2,63	2,59	2,58	2,56	2,54	2,51	2,49	2,47
	8	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56	2,54	2,50	2,46	2,42	2,40	2,38	2,36	2,34	2,32	2,29
	9	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44	2,42	2,38	2,34	2,30	2,28	2,25	2,23	2,21	2,18	2,16
	10	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35	2,32	2,28	2,24	2,20	2,18	2,16	2,13	2,11	2,08	2,06
	11	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27	2,25	2,21	2,17	2,12	2,10	2,08	2,05	2,03	2,00	1,97
	12	3,18	2,81	2,61	2,48	2,39	2,33	2,28	2,24	2,21	2,19	2,15	2,10	2,06	2,04	2,01	1,99	1,96	1,93	1,90
	13	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,20	2,16	2,14	2,10	2,05	2,01	1,98	1,96	1,93	1,90	1,88	1,85
	14	3,10	2,73	2,52	2,39	2,31	2,24	2,19	2,15	2,12	2,10	2,05	2,01	1,96	1,94	1,91	1,89	1,86	1,83	1,80
	15	3,07	2,70	2,49	2,36	2,27	2,21	2,16	2,12	2,09	2,06	2,02	1,97	1,92	1,90	1,87	1,85	1,82	1,79	1,76
	16	3,05	2,67	2,46	2,33	2,24	2,18	2,13	2,09	2,06	2,03	1,99	1,94	1,89	1,87	1,84	1,81	1,78	1,75	1,72
	17	3,03	2,64	2,44	2,31	2,22	2,15	2,10	2,06	2,03	2,00	1,96	1,91	1,86	1,84	1,81	1,78	1,75	1,72	1,69
	18	3,01	2,62	2,42	2,29	2,20	2,13	2,08	2,04	2,00	1,98	1,93	1,89	1,84	1,81	1,78	1,75	1,72	1,69	1,66
	19	2,99	2,61	2,40	2,27	2,18	2,11	2,06	2,02	1,98	1,96	1,91	1,86	1,81	1,79	1,76	1,73	1,70	1,67	1,63
	20	2,97	2,59	2,38	2,25	2,16	2,09	2,04	2,00	1,96	1,94	1,89	1,84	1,79	1,77	1,74	1,71	1,68	1,64	1,61
	21	2,96	2,57	2,36	2,23	2,14	2,08	2,02	1,98	1,95	1,92	1,88	1,83	1,78	1,75	1,72	1,69	1,66	1,62	1,59
	22	2,95	2,56	2,35	2,22	2,13	2,06	2,01	1,97	1,93	1,90	1,86	1,81	1,76	1,73	1,70	1,67	1,64	1,60	1,57
	23	2,94	2,55	2,34	2,21	2,11	2,05	1,99	1,95	1,92	1,89	1,84	1,80	1,74	1,72	1,69	1,66	1,62	1,59	1,55
	24	2,93	2,54	2,33	2,19	2,10	2,04	1,98	1,94	1,91	1,88	1,83	1,78	1,73	1,70	1,67	1,64	1,61	1,57	1,53
	25	2,92	2,53	2,32	2,18	2,09	2,02	1,97	1,93	1,89	1,87	1,82	1,77	1,72	1,69	1,66	1,63	1,59	1,56	1,52
	26	2,91	2,52	2,31	2,17	2,08	2,01	1,96	1,92	1,88	1,86	1,81	1,76	1,71	1,68	1,65	1,61	1,58	1,54	1,50
	27	2,90	2,51	2,30	2,17	2,07	2,00	1,95	1,91	1,87	1,85	1,80	1,75	1,70	1,67	1,64	1,60	1,57	1,53	1,49
	28	2,89	2,50	2,29	2,16	2,06	2,00	1,94	1,90	1,87	1,84	1,79	1,74	1,69	1,66	1,63	1,59	1,56	1,52	1,48
	29	2,89	2,50	2,28	2,15	2,06	1,99	1,93	1,89	1,86	1,83	1,78	1,73	1,68	1,65	1,62	1,58	1,55	1,51	1,47
	30	2,88	2,49	2,28	2,14	2,05	1,98	1,93	1,88	1,85	1,82	1,77	1,72	1,67	1,64	1,61	1,57	1,54	1,50	1,46
40	2,84	2,44	2,23	2,09	2,00	1,93	1,87	1,83	1,79	1,76	1,71	1,66	1,61	1,57	1,54	1,51	1,47	1,42	1,38	
60	2,79	2,39	2,18	2,04	1,95	1,87	1,82	1,77	1,74	1,71	1,66	1,60	1,54	1,51	1,48	1,44	1,40	1,35	1,29	
120	2,75	2,35	2,13	1,99	1,90	1,82	1,77	1,72	1,68	1,65	1,60	1,55	1,48	1,45	1,41	1,37	1,32	1,26	1,19	
:	2,71	2,30	2,08	1,94	1,85	1,77	1,72	1,67	1,63	1,60	1,55	1,49	1,42	1,38	1,34	1,30	1,24	1,17	1,00	

## A $\chi^2$ -eloszlás táblázata

<i>f</i>	Valószínűségek											
	0,99	0,95	0,90	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,01	0,001
1	0,03157	0,00393	0,0158	0,0642	0,148	0,455	1,074	1,642	2,706	3,841	6,635	10,827
2	0,0201	0,103	0,211	0,446	0,713	1,386	2,408	3,219	4,605	5,991	9,210	13,815
3	0,115	0,352	0,584	1,005	1,424	2,366	3,665	4,642	6,251	7,815	11,345	16,266
4	0,297	0,711	1,064	1,649	2,195	3,357	4,878	5,989	7,779	9,488	13,277	18,467
5	0,554	1,145	1,610	2,343	3,000	4,351	6,064	7,289	9,236	11,070	15,086	20,515
6	0,872	1,635	2,204	3,070	3,828	5,348	7,231	8,558	10,645	12,592	16,812	22,457
7	1,239	2,167	2,833	3,822	4,671	6,346	8,383	9,803	12,017	14,067	18,475	24,322
8	1,646	2,733	3,490	4,594	5,527	7,344	9,524	11,030	13,362	15,507	20,090	26,125
9	2,088	3,325	4,168	5,380	6,393	8,343	10,656	12,242	14,684	16,919	21,666	27,877
10	2,558	3,940	4,865	6,179	7,267	9,342	11,781	13,442	15,987	18,307	23,209	29,588
11	3,053	4,575	5,578	6,989	8,148	10,341	12,899	14,631	17,275	19,675	24,725	31,264
12	3,571	5,226	6,304	7,807	9,034	11,340	14,011	15,812	18,549	21,026	26,217	32,909
13	4,107	5,892	7,042	8,634	9,926	12,340	15,119	16,985	19,812	22,362	27,688	34,528
14	4,660	6,571	7,790	9,467	10,821	13,339	16,222	18,151	21,064	23,685	29,141	36,123
15	5,229	7,261	8,547	10,307	11,721	14,339	17,322	19,311	22,307	24,996	30,578	37,697
16	5,812	7,962	9,312	11,152	11,624	15,338	18,418	20,465	23,542	26,296	32,000	39,252
17	6,408	8,672	10,085	12,002	13,531	16,338	19,511	21,615	24,769	27,587	33,409	40,790
18	7,015	9,390	10,865	12,857	14,440	17,338	20,601	22,760	25,989	28,869	34,805	42,312
19	7,633	10,117	11,651	13,716	15,352	18,338	21,689	23,900	27,204	30,144	36,191	43,820
20	8,260	10,851	12,443	14,578	16,266	19,337	22,775	25,038	28,412	31,410	37,566	45,315
21	8,897	11,591	13,240	15,445	17,182	20,337	23,858	26,171	29,615	32,671	38,932	46,797
22	9,542	12,338	14,041	16,314	18,101	21,337	24,939	27,301	30,813	33,924	40,289	48,268
23	10,196	13,091	14,848	17,187	19,021	22,337	26,018	28,429	32,007	35,172	41,638	49,728
24	10,856	13,848	15,659	18,062	19,943	23,337	27,096	29,553	33,196	36,415	42,980	51,179
25	11,524	14,611	16,473	18,940	20,867	24,337	28,172	30,675	34,382	37,652	44,314	52,620
26	12,198	15,379	17,292	19,820	21,792	25,336	29,246	31,795	35,563	38,885	45,642	54,052
27	12,879	16,151	18,114	20,703	22,719	26,336	30,319	32,912	36,741	40,113	46,963	55,476
28	13,565	16,928	18,939	21,588	23,647	27,336	31,391	34,027	37,916	41,337	48,278	56,893
29	14,256	17,708	19,768	22,475	24,577	28,336	32,461	35,139	39,087	42,557	49,588	58,302
30	14,953	18,493	20,599	23,364	25,508	29,336	33,530	36,250	40,256	43,773	50,892	59,703
32	16,362	20,072	22,271	25,148	27,373	31,336	35,665	38,466	42,585	46,194	53,486	62,487
34	17,789	21,664	23,952	26,938	29,242	33,336	37,795	40,676	44,903	48,602	56,061	65,247
36	19,233	23,269	25,643	28,735	31,115	35,336	39,922	42,879	47,212	50,999	58,619	67,985
38	20,691	24,884	27,343	30,537	32,992	37,335	42,045	45,076	49,513	53,384	61,162	70,703
40	22,164	26,509	29,051	32,345	34,872	39,335	44,165	47,269	51,805	55,759	63,691	73,402
42	23,650	28,144	30,765	34,157	36,755	41,335	46,282	49,456	54,090	58,124	66,206	76,084
44	25,148	29,787	32,487	35,974	38,641	43,335	48,396	51,639	56,369	60,481	68,710	78,750
46	26,657	31,439	34,215	37,795	40,529	45,335	50,507	53,818	58,641	62,830	71,201	81,400
48	28,177	33,098	35,949	39,621	42,420	47,335	52,616	55,993	60,907	65,171	73,683	84,037
50	29,707	34,764	37,689	41,449	44,313	49,335	54,723	58,164	63,167	67,505	76,154	86,661
52	31,246	36,437	39,433	43,281	46,209	51,335	56,827	60,332	65,422	69,832	78,616	89,272
54	32,793	38,116	41,183	45,117	48,106	53,335	58,930	62,496	67,673	72,153	81,009	91,872
56	34,350	39,801	42,937	46,955	50,005	55,335	61,031	64,658	69,919	74,468	83,513	94,461
58	35,913	41,492	44,696	48,797	51,906	57,335	63,129	66,816	72,160	76,778	85,950	97,039
60	37,485	43,188	46,459	50,641	53,809	59,335	65,227	68,972	74,397	79,082	88,379	99,607
62	39,063	44,889	48,226	52,487	55,714	61,335	67,322	71,125	76,630	81,381	90,802	102,166
64	40,649	46,595	49,996	54,336	57,620	63,335	69,416	73,276	78,860	83,675	93,217	104,716
66	42,240	48,305	51,770	56,188	59,527	65,335	71,508	75,424	81,085	85,965	95,626	107,258
68	43,838	50,020	53,548	58,042	61,436	67,335	73,600	77,571	83,308	88,250	98,028	109,791
70	45,442	51,739	55,329	59,898	63,346	69,334	75,689	79,715	85,527	90,531	100,425	112,317

A standard normális eloszlású valószínűségi változó  
eloszlásfüggvényének táblázata

$$[\Phi(-x) = 1 - \Phi(x)]$$

0,00	0,5000	0,43	0,6664	0,86	0,8051	1,29	0,9015	1,72	0,9572	2,30	0,9893
0,01	0,5040	0,44	0,6700	0,87	0,8078	1,30	0,9032	1,73	0,9582	2,32	0,9898
0,02	0,5080	0,45	0,6736	0,88	0,8106	1,31	0,9049	1,74	0,9591	2,34	0,9904
0,03	0,5120	0,46	0,6772	0,89	0,8133	1,32	0,9066	1,75	0,9599	2,36	0,9909
0,04	0,5160	0,47	0,6808	0,90	0,8159	1,33	0,9082	1,76	0,9608	2,38	0,9913
0,05	0,5199	0,48	0,6844	0,91	0,8186	1,34	0,9099	1,77	0,9616	2,40	0,9918
0,06	0,5239	0,49	0,6879	0,92	0,8212	1,35	0,9115	1,78	0,9625	2,42	0,9922
0,07	0,5279	0,50	0,6915	0,93	0,8238	1,36	0,9131	1,79	0,9633	2,44	0,9927
0,08	0,5319	0,51	0,6950	0,94	0,8264	1,37	0,9147	1,80	0,9641	2,46	0,9931
0,09	0,5359	0,52	0,6985	0,95	0,8289	1,38	0,9162	1,81	0,9649	2,48	0,9934
0,10	0,5398	0,53	0,7019	0,96	0,8315	1,39	0,9177	1,82	0,9656	2,50	0,9938
0,11	0,5438	0,54	0,7054	0,97	0,8340	1,40	0,9192	1,83	0,9664	2,52	0,9941
0,12	0,5478	0,55	0,7088	0,98	0,8365	1,41	0,9207	1,84	0,9671	2,54	0,9945
0,13	0,5517	0,56	0,7123	0,99	0,8389	1,42	0,9222	1,85	0,9678	2,56	0,9948
0,14	0,5557	0,57	0,7157	1,00	0,8413	1,43	0,9236	1,86	0,9686	2,58	0,9951
0,15	0,5596	0,58	0,7190	1,01	0,8438	1,44	0,9251	1,87	0,9693	2,60	0,9953
0,16	0,5636	0,59	0,7224	1,02	0,8461	1,45	0,9265	1,88	0,9699	2,62	0,9956
0,17	0,5675	0,60	0,7257	1,03	0,8485	1,46	0,9279	1,89	0,9706	2,64	0,9959
0,18	0,5714	0,61	0,7291	1,04	0,8508	1,47	0,9292	1,90	0,9713	2,66	0,9961
0,19	0,5753	0,62	0,7324	1,05	0,8531	1,48	0,9306	1,91	0,9719	2,68	0,9963
0,20	0,5793	0,63	0,7357	1,06	0,8554	1,49	0,9319	1,92	0,9726	2,70	0,9965
0,21	0,5832	0,64	0,7389	1,07	0,8577	1,50	0,9332	1,93	0,9732	2,72	0,9967
0,22	0,5871	0,65	0,7422	1,08	0,8599	1,51	0,9345	1,94	0,9738	2,74	0,9969
0,23	0,5910	0,66	0,7454	1,09	0,8621	1,52	0,9357	1,95	0,9744	2,76	0,9971
0,24	0,5948	0,67	0,7486	1,10	0,8643	1,53	0,9370	1,96	0,9750	2,78	0,9973
0,25	0,5987	0,68	0,7517	1,11	0,8665	1,54	0,9382	1,97	0,9756	2,80	0,9974
0,26	0,6026	0,69	0,7549	1,12	0,8686	1,55	0,9394	1,98	0,9761	2,82	0,9976
0,27	0,6064	0,70	0,7580	1,13	0,8708	1,56	0,9406	1,99	0,9767	2,84	0,9977
0,28	0,6103	0,71	0,7611	1,14	0,8729	1,57	0,9418	2,00	0,9772	2,86	0,9979
0,29	0,6141	0,72	0,7642	1,15	0,8749	1,58	0,9429	2,02	0,9778	2,88	0,9980
0,30	0,6179	0,73	0,7673	1,16	0,8770	1,59	0,9441	2,04	0,9793	2,90	0,9981
0,31	0,6217	0,74	0,7703	1,17	0,8790	1,60	0,9452	2,06	0,9803	2,92	0,9982
0,32	0,6255	0,75	0,7734	1,18	0,8810	1,61	0,9463	2,08	0,9812	2,94	0,9984
0,33	0,6293	0,76	0,7764	1,19	0,8830	1,62	0,9474	2,10	0,9821	2,96	0,9985
0,34	0,6331	0,77	0,7794	1,20	0,8849	1,63	0,9484	2,12	0,9830	2,98	0,9986
0,35	0,6368	0,78	0,7823	1,21	0,8869	1,64	0,9495	2,14	0,9838	3,00	0,9987
0,36	0,6406	0,79	0,7853	1,22	0,8888	1,65	0,9505	2,16	0,9846	3,20	0,9993
0,37	0,6443	0,80	0,7881	1,23	0,8907	1,66	0,9515	2,18	0,9854	3,40	0,9996
0,38	0,6480	0,81	0,7910	1,24	0,8925	1,67	0,9525	2,20	0,9861	3,60	0,9998
0,39	0,6517	0,82	0,7939	1,25	0,8944	1,68	0,9535	2,22	0,9868	3,80	0,9999
0,40	0,6554	0,83	0,7967	1,26	0,8962	1,69	0,9545	2,24	0,9875		
0,41	0,6591	0,84	0,7995	1,27	0,8980	1,70	0,9554	2,26	0,9881		
0,42	0,6628	0,85	0,8023	1,28	0,8997	1,71	0,9564	2,28	0,9887		



A korreláció szignifikancia szintjének a kritikus értékei

<i>szf</i>	80%	90%	95%	98%	99%
1	0,9511	0,9877	0,9969	0,9995	0,9999
2	0,8000	0,9000	0,9500	0,9800	0,9900
3	0,6870	0,8054	0,8783	0,9343	0,9587
4	0,6084	0,7293	0,8114	0,8822	0,9172
5	0,5509	0,6694	0,7545	0,8329	0,8745
6	0,5067	0,6215	0,7067	0,7887	0,8343
7	0,4716	0,5822	0,6664	0,7498	0,7977
8	0,4428	0,5494	0,6319	0,7155	0,7646
9	0,4187	0,5214	0,6021	0,6851	0,7348
10	0,3981	0,4973	0,5760	0,6581	0,7079
11	0,3802	0,4762	0,5529	0,6339	0,6835
12	0,3646	0,4575	0,5324	0,6120	0,6614
13	0,3507	0,4409	0,5140	0,5923	0,6411
14	0,3383	0,4259	0,4973	0,5742	0,6226
15	0,3271	0,4124	0,4821	0,5577	0,6055
16	0,3170	0,4000	0,4683	0,5425	0,5897
17	0,3077	0,3887	0,4555	0,5285	0,5751
18	0,2992	0,3783	0,4438	0,5155	0,5614
19	0,2914	0,3687	0,4329	0,5034	0,5487
20	0,2841	0,3598	0,4227	0,4921	0,5368
21	0,2774	0,3515	0,4132	0,4815	0,5256
22	0,2711	0,3438	0,4044	0,4716	0,5151
23	0,2653	0,3365	0,3961	0,4622	0,5052
24	0,2598	0,3297	0,3882	0,4534	0,4958
25	0,2546	0,3233	0,3809	0,4451	0,4869
26	0,2497	0,3172	0,3739	0,4372	0,4785
27	0,2451	0,3115	0,3673	0,4297	0,4705
28	0,2407	0,3061	0,3610	0,4226	0,4629
29	0,2366	0,3009	0,3550	0,4158	0,4556
30	0,2327	0,2960	0,3494	0,4093	0,4487
32	0,2254	0,2869	0,3388	0,3972	0,4357
34	0,2187	0,2785	0,3291	0,3862	0,4238
36	0,2126	0,2709	0,3202	0,3760	0,4128
38	0,2070	0,2638	0,3120	0,3665	0,4026
40	0,2018	0,2573	0,3044	0,3578	0,3932
45	0,1903	0,2429	0,2876	0,3384	0,3721
50	0,1806	0,2306	0,2732	0,3218	0,3542
60	0,1650	0,2108	0,2500	0,2948	0,3248
70	0,1528	0,1954	0,2319	0,2737	0,3017
80	0,1430	0,1829	0,2172	0,2565	0,2830
90	0,1348	0,1726	0,2050	0,2422	0,2673
100	0,1279	0,1638	0,1946	0,2301	0,2540
120	0,1168	0,1496	0,1779	0,2104	0,2324