TÉRSTATISZTIKA

Dr. Sipos Tibor
Dr. Török Ádám
Szabó Zsombor

MŰEGYETEM 1782

KUKG

BUDAPESTI MŰSZAKI ÉS GAZDASÁGTUDOMÁNYI EGYETEM
Közlekedésmérnöki és Járműmérnöki Kar

Tartalom

Térstatisztika

Térbeli interakciók

Gravitációs modell

Tartalom

Térstatisztika

↓

Térbeli interakciók

↓

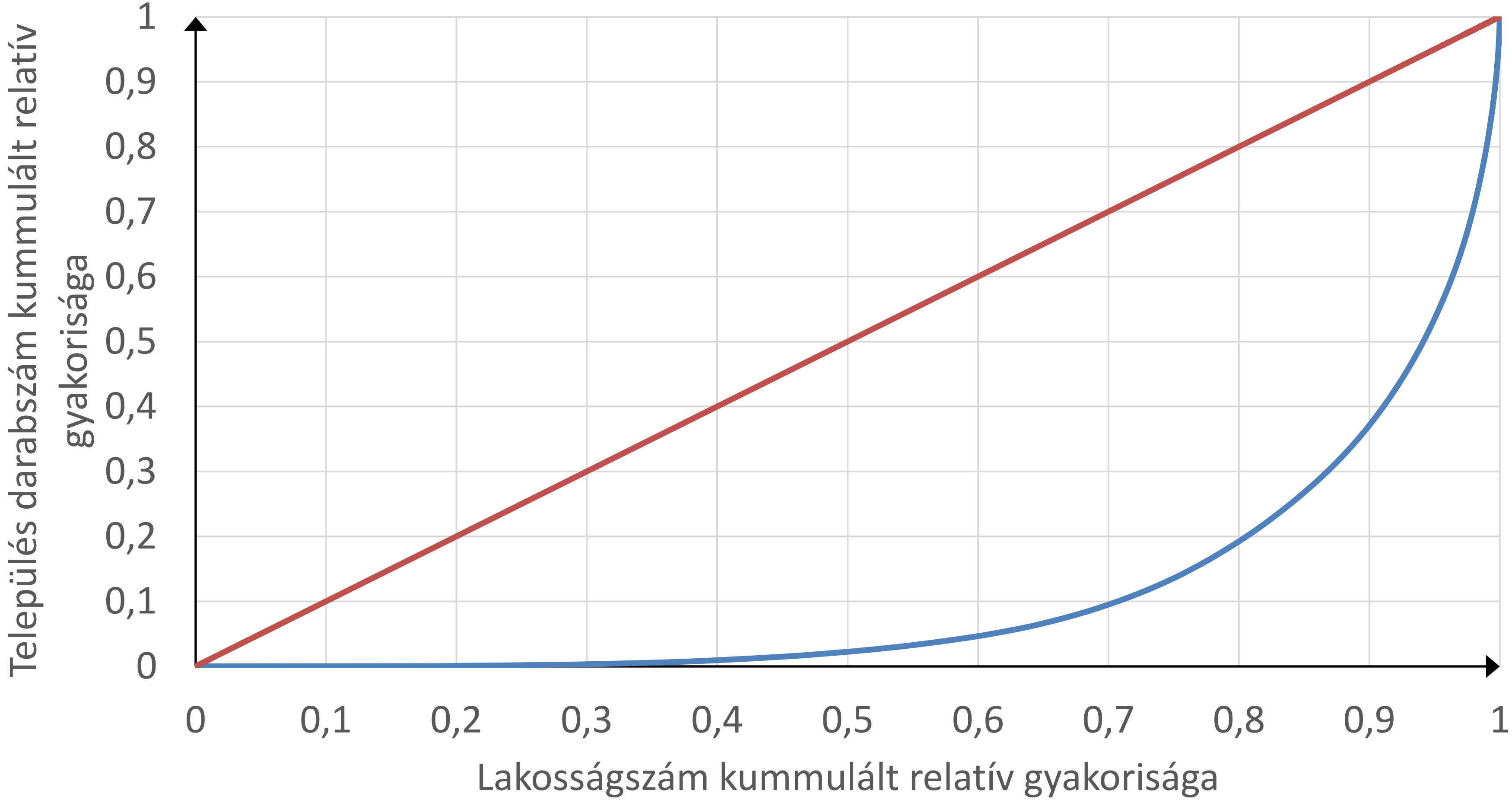
Gravitációs modell

Térbeli egységek statisztikai elemzése

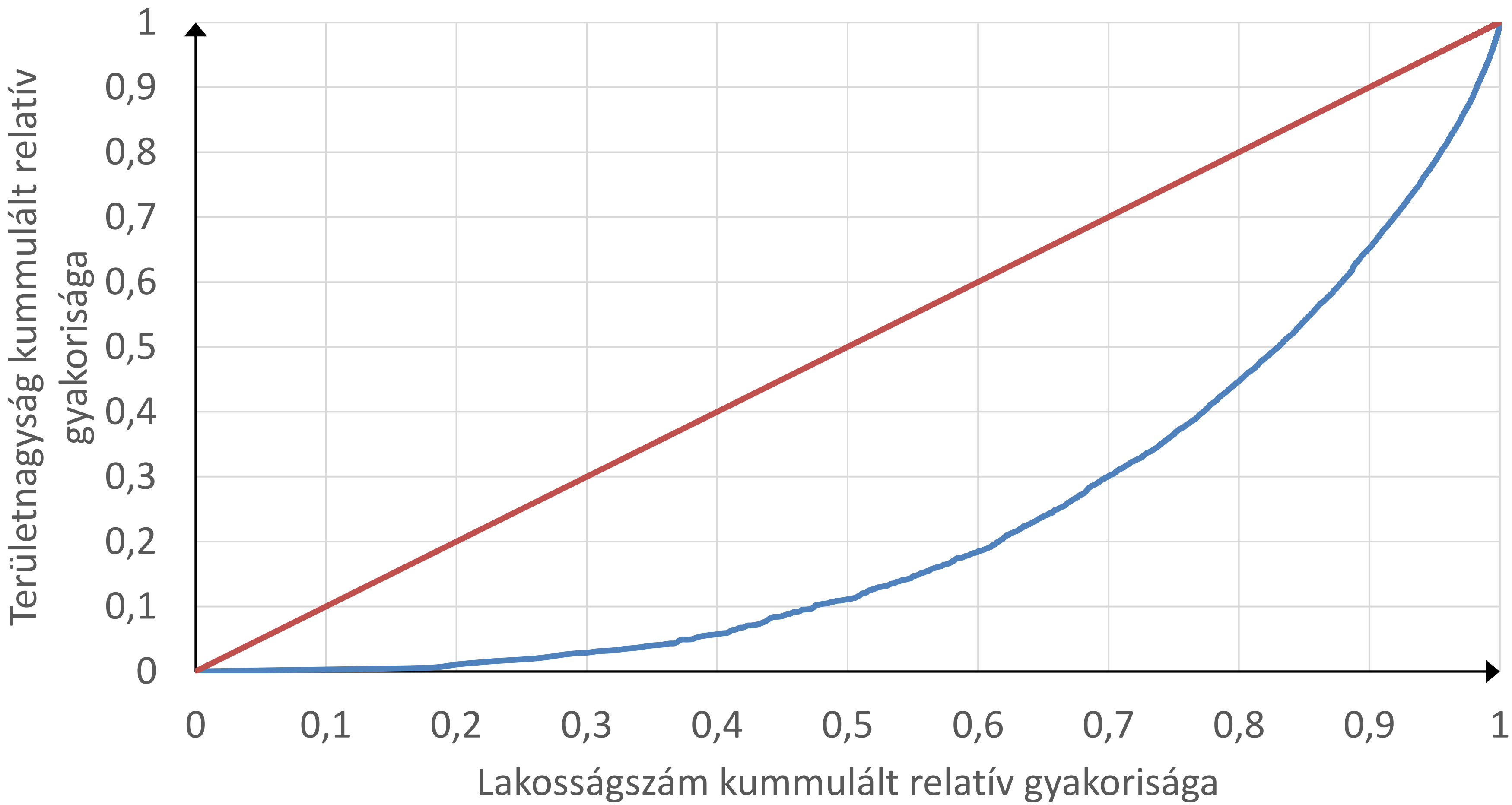
- A térbeli egységekre ugyanúgy alkalmazhatók a statisztikából megismert módszerek
- Legtöbbször azonban ezen adatok nem normális eloszlást követnek
- Normáleloszlás esetén $\bar{x} = E(x) = x_{MO} = x_{ME}$

Magyarország lakosságszáma	
Várható érték	3099,3252
Standard hiba	576,025573
Medián	799
Módusz	183
Szórás	32355,0031
Minta varianciája	1046846228
Csúcsosság	2695,92513
Ferdeség	50,1646485
Tartomány	1749726
Minimum	8
Maximum	1749734
Összeg	9778371
Darabszám	3155

Lorentz görbe egységnyi település mellett



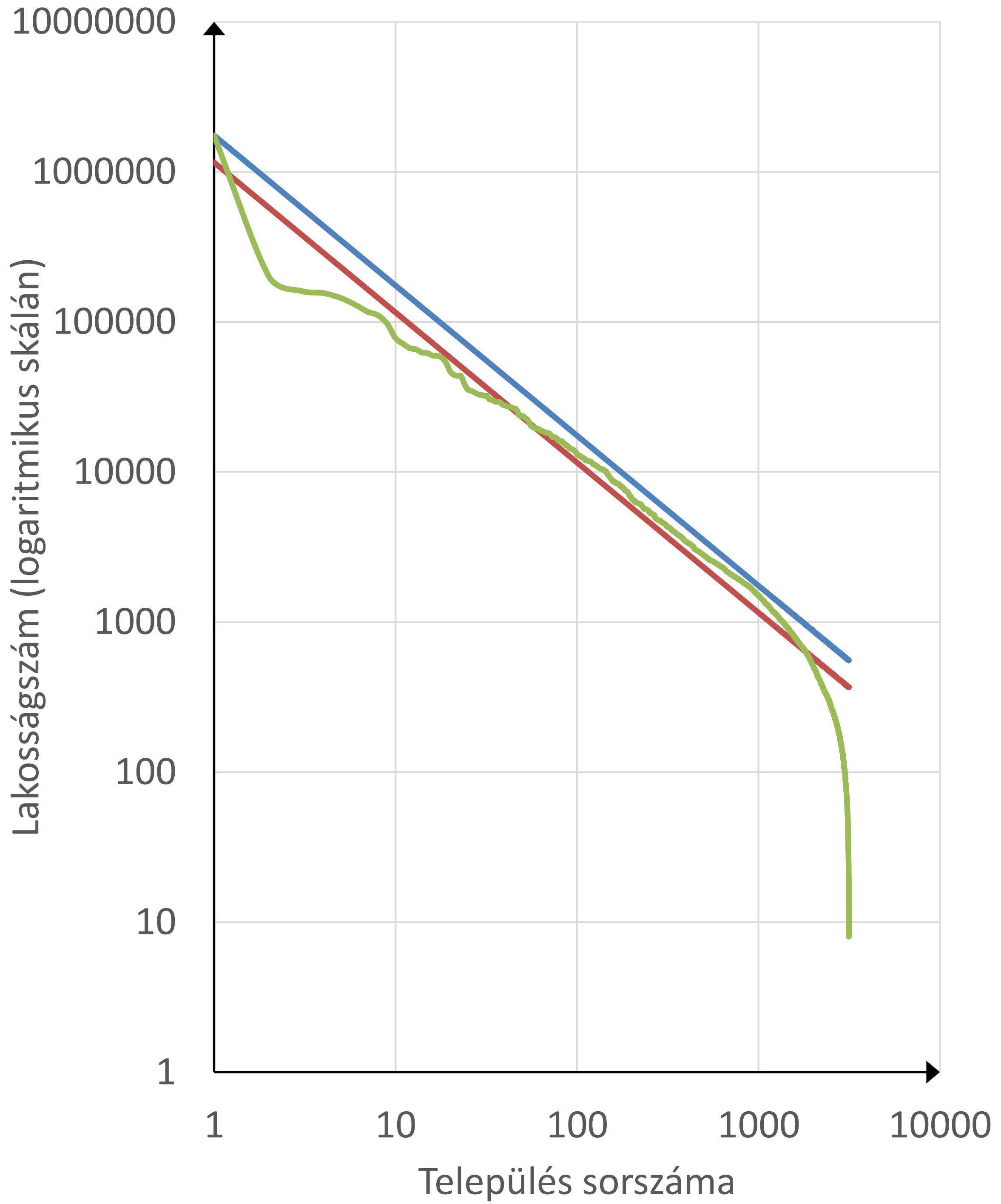
Lorentz görbe terület alapon



Lakosságszám

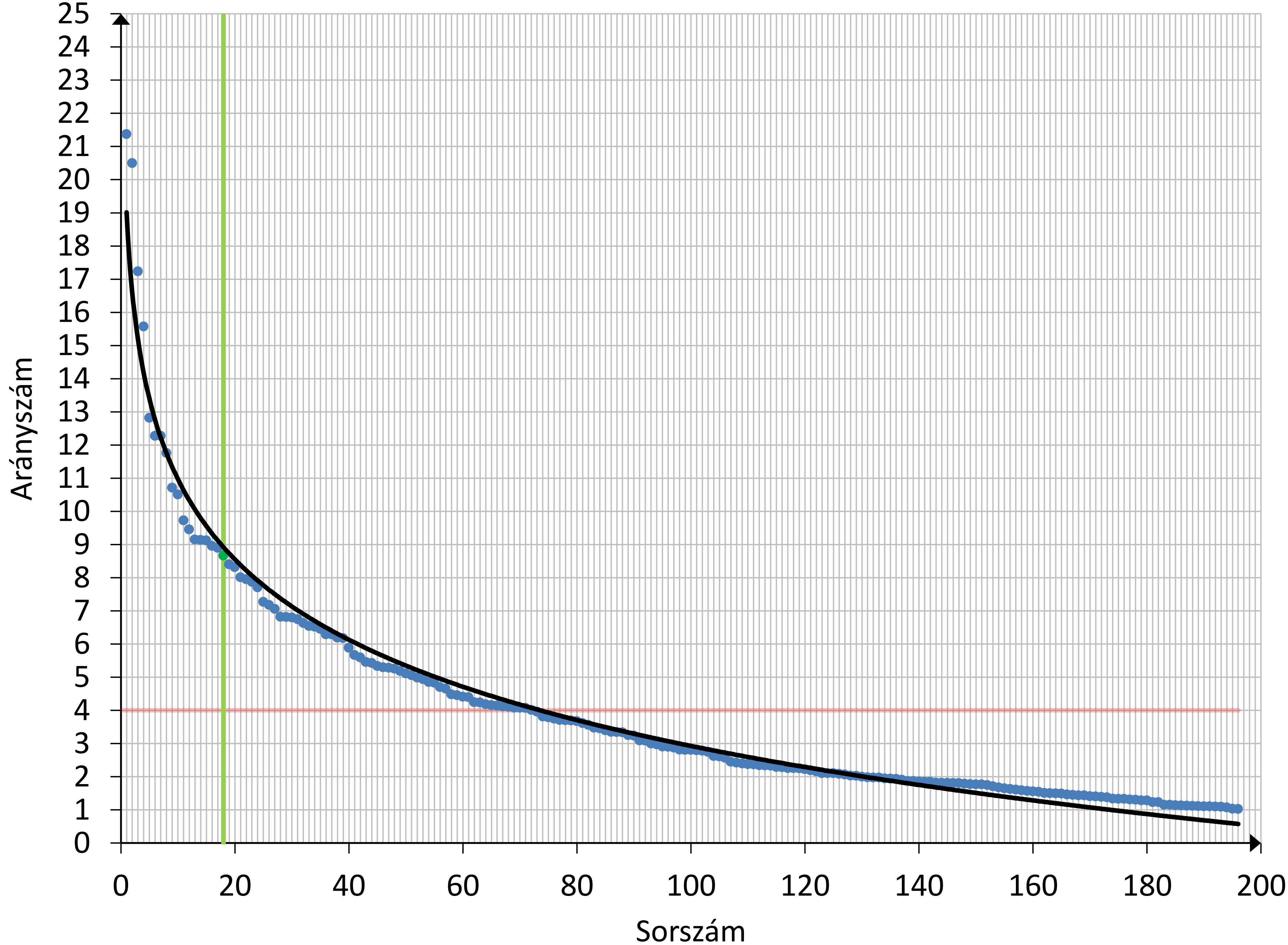
- Felix Auerbach (1856-1933) német fizikus (1917)
  - Egy adott területen ideális esetben ha sorba rendezzük lakosságszám szerint a településeket, akkor az n. település lakosságszáma pontosan $1/n$ része a legnagyobbnak
- Ez a Zipf-eloszlás egy kitüntetett esete
- Később bebizonyosodott, hogy az $1/n^{1,07}$ jobban leírja a települések viszonyát

- Különböző városstruktúrák
  - Auberbach tétele alapján jelenlegi budapesti lakosságszámmal
  - Auberbach tétele alapján 10000000 összlakossággal
  - Jelenlegi magyar helyzet

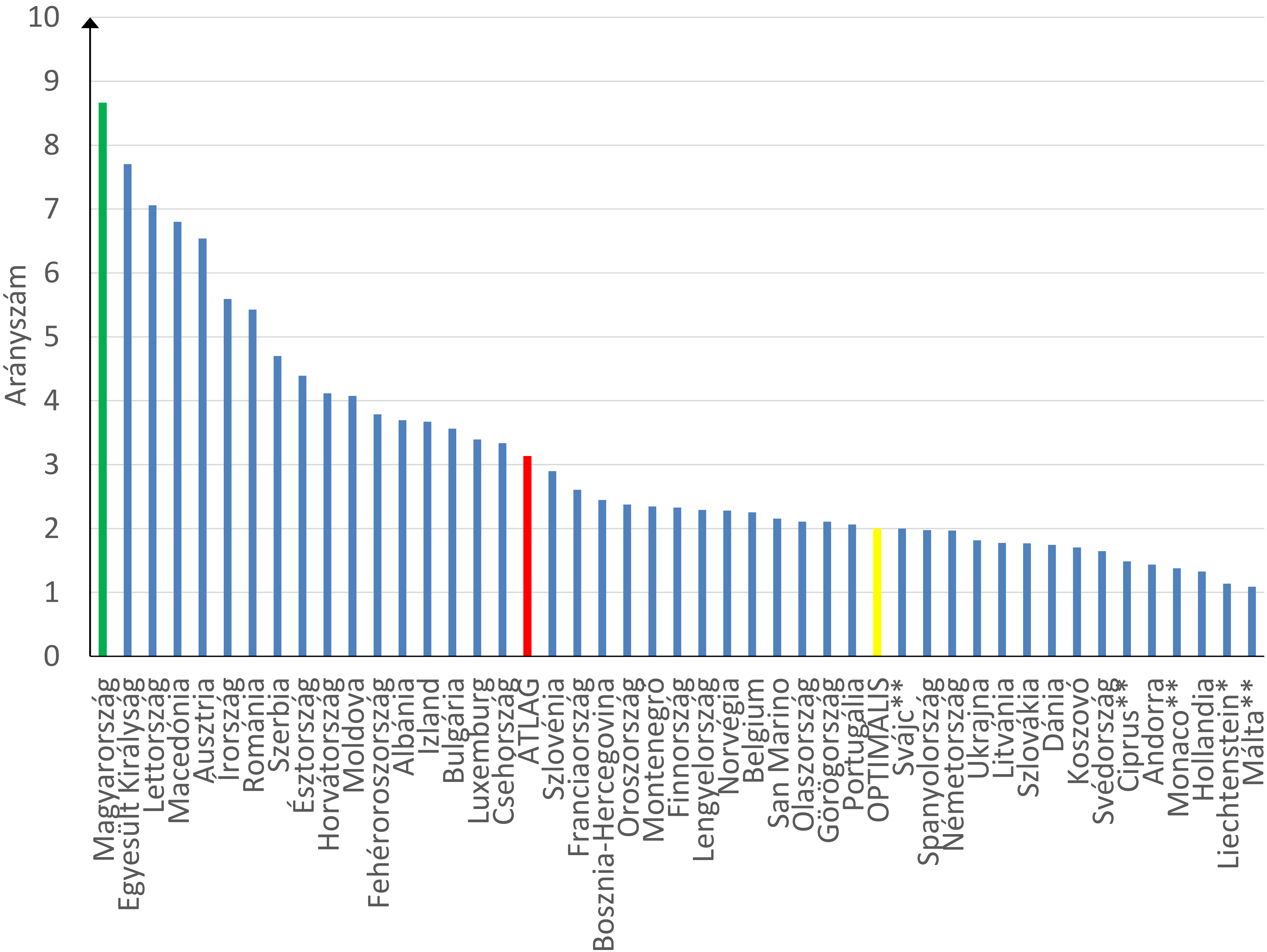


Auerbach tétele

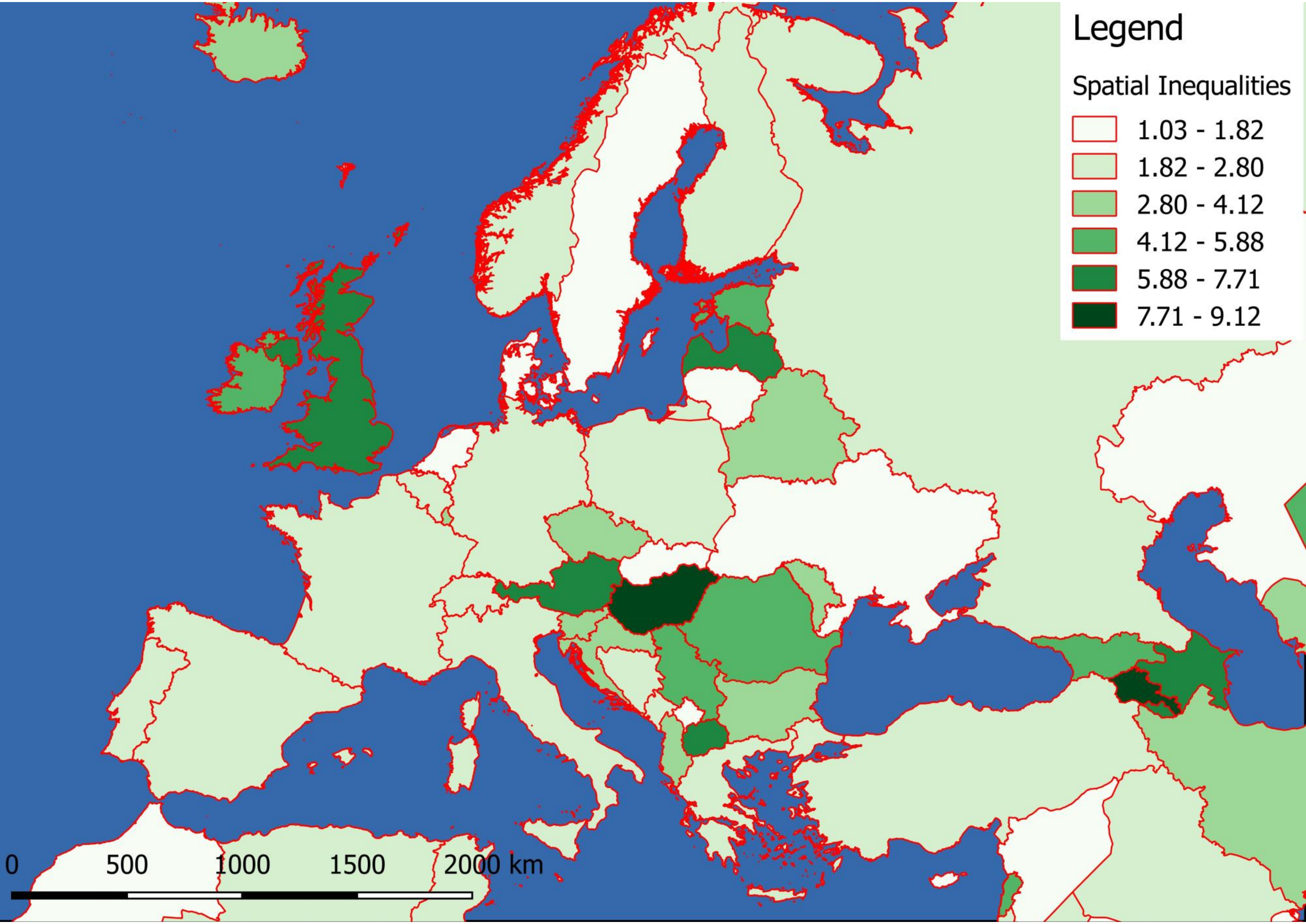
- Auerbach tétele alapján definiálható egy viszonyszám, amely a legnagyobb és a második legnagyobb település lakosságszámának az aránya (nevezzük egyszerűen arányszámnak)
- Ez mintegy „vízfejségi mutató” jelenhetne meg egy adott területre

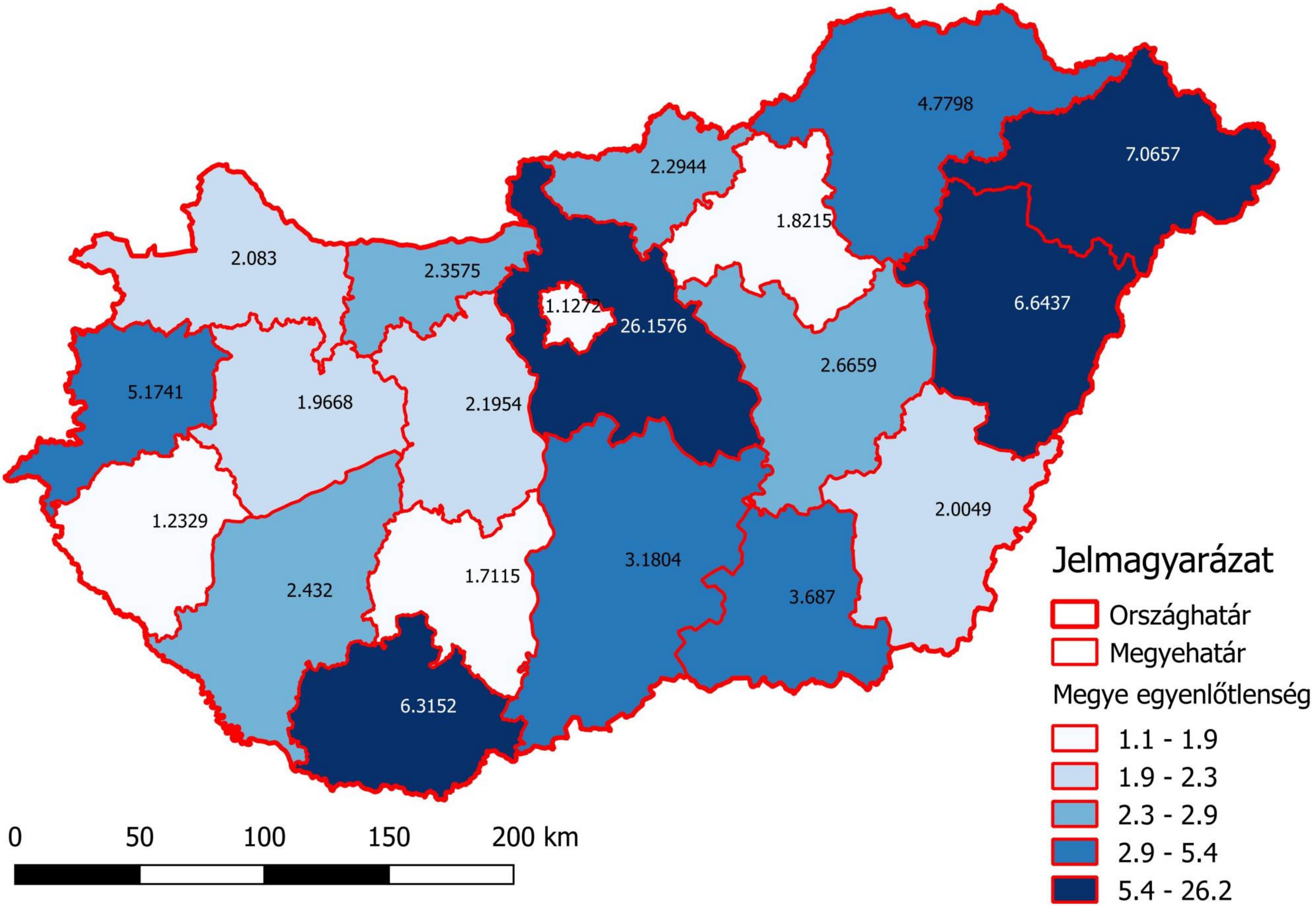


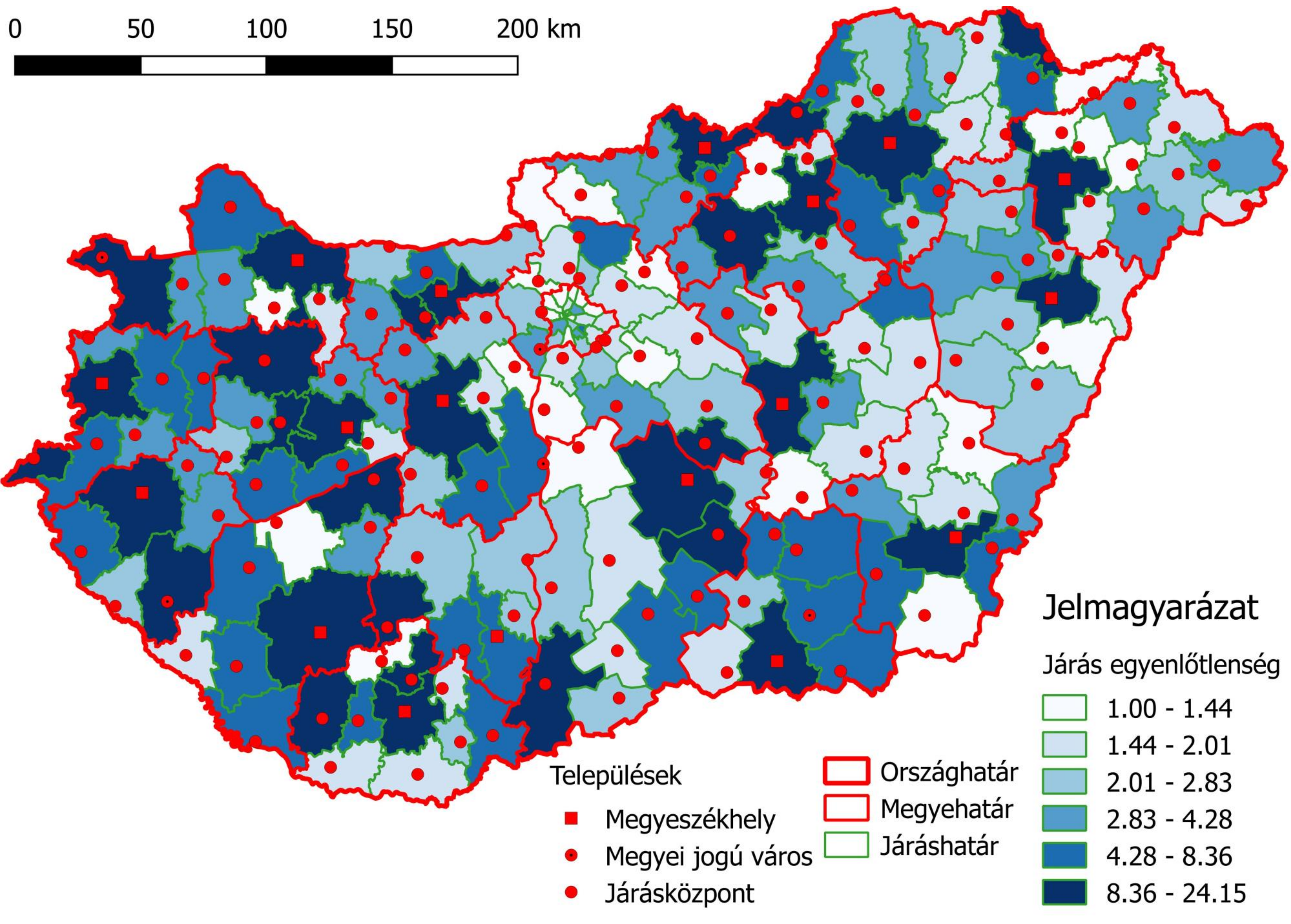
Arányszám
Sorszám
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
0
20
40
60
80
100
120
140
160
180
200



Arányszám
10
9
8
7
6
5
4
3
2
1
0
Magyarország
Egyesült Királyság
Lettország
Macedónia
Ausztria
Írország
Románia
Szerbia
Észtország
Horvátország
Moldova
Fehéroroszország
Albánia
Izland
Bulgária
Luxemburg
Csehország
ÁTLAG
Szlovénia
Franciaország
Bosznia-Hercegovina
Oroszország
Montenegró
Finnország
Lengyelország
Norvégia
Belgium
San Marino
Olaszország
Görögország
Portugália
OPTIMÁLIS
Svájc**
Spanyolország
Németország
Ukrajna
Litvánia
Szlovákia
Dánia
Koszovó
Svédország
Ciprus**
Andorra
Monaco**
Hollandia
Liechtenstein*
Málta**


Legend
Spatial Inequalities
1.03 - 1.82
1.82 - 2.80
2.80 - 4.12
4.12 - 5.88
5.88 - 7.71
7.71 - 9.12
0
500
1000
1500
2000 km


4.7798
7.0657
2.2944
1.8215
2.083
2.3575
6.6437
1.1272
26.1576
2.6659
5.1741
1.9668
2.1954
2.0049
1.2329
3.1804
1.7115
2.432
3.687
6.3152
Jelmagyarázat
Országhatár
Megyehatár
Megye egyenlőtlenség
1.1 - 1.9
1.9 - 2.3
2.3 - 2.9
2.9 - 5.4
5.4 - 26.2
0
50
100
150
200 km



0
50
100
150
200 km
Jelmagyarázat
Járás egyenlőtlenség
1.00 - 1.44
1.44 - 2.01
2.01 - 2.83
2.83 - 4.28
4.28 - 8.36
8.36 - 24.15
Települések
Megyeszékhely
Megyei jogú város
Járásközpont
Országhatár
Megyehatár
Járáshatár

Nem normális eloszlás következményei

- u-próba -> Mann-Whitney próba
- t-próba -> Wilcoxon-próba
- F-próba -> Levene-próba

Térbeli koncentráció és szegregáció

- A térbeli egységek esetében fontos szerepe van a különböző koncentrációt és szegregációt mérő indexeknek
- Koncentráció: területi eloszlások egyenlőtlenségeit jelzik
  - Hirschmann–Herfindahl-index (Simpson-index)
  - Gini-mutató
  - Theil-index
- Szegregáció: két területi eloszlás különbözőségét mérik
  - Hoover-index

Simpson-index

$$HI = \sum_{i=1}^{n} \left(\frac{y_i}{\sum y_i} \right)^2$$

- Értéke:
  - $\frac{1}{n}$ a teljes egyenlőség esetén
  - 1 a tökéletes koncentráltság esetén
- Mivel az értéke függ n értékétől ezért létezik normalizált alakja

$$HI^* = \frac{HI - \frac{1}{n}}{1 - \frac{1}{n}} \in [0;1]$$

Gini-mutató

$$G = 1 - \sum\left[\left(f'_{r_i} - f'_{r_{i-1}}\right)\left(Z'_{r_i} + Z'_{r_{i-1}}\right)\right]$$

- $G \in [0..1]$
  - $G = 0$ – teljes egyenlőség
  - $G = 1$ – teljes egyenlőtlenség
  - $G \geq 0{,}4$ – jelentősebb egyenlőtlenség

Theil-index

- A Theil-index az általános entrópia index egy speciális esete

$$T = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i}{\bar{y}}\ln\frac{y_i}{\bar{y}}\right)$$

- Ha $T = 0$, ha teljes egyenlőség van
- Ha $T = \ln n$, akkor tökéletes koncentráltság áll fenn

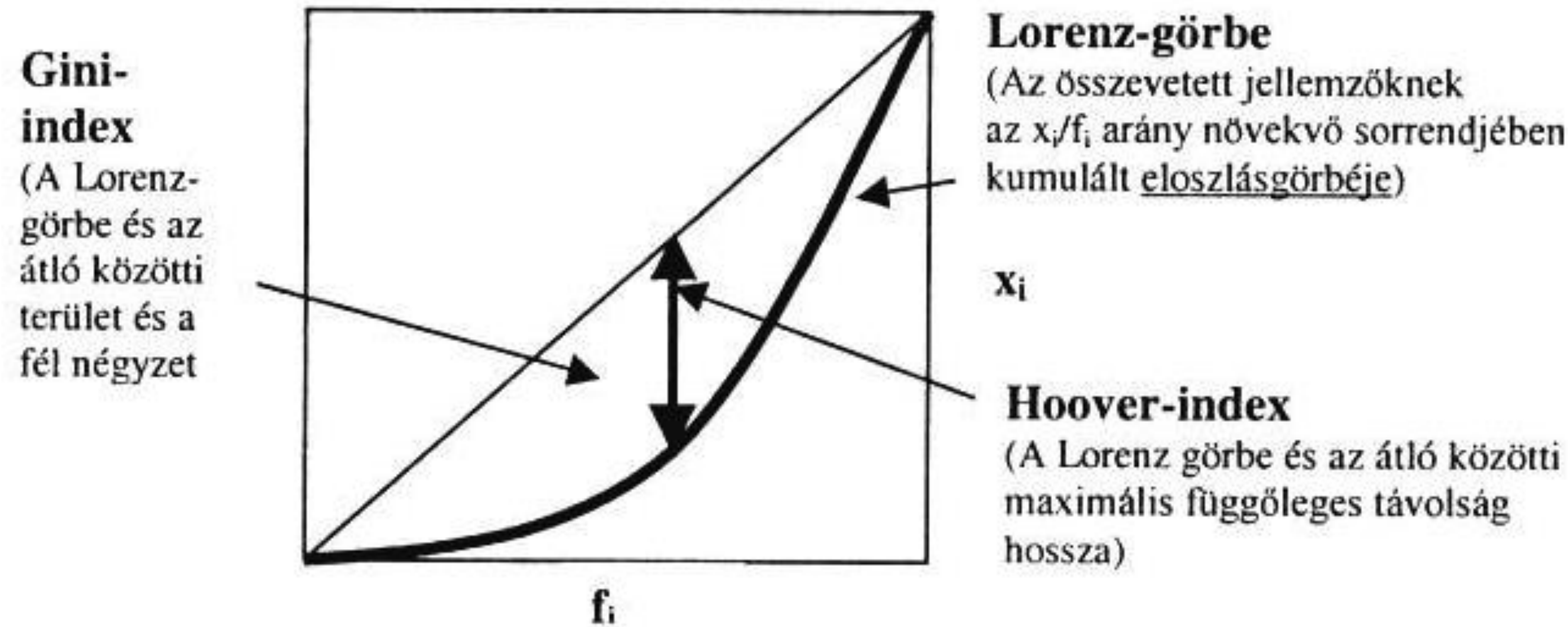
Hoover-index

- Számos neve van
  - Schutz-index
  - Robin Hood-index (Rózsa Sándor-index)
  - Duncan-index
  - Krugman-index (2-vel osztás nélkül)
- Jelentése: az y ismérv hány százalékát kell átcsoportosítani, hogy megegyezzen az x területegységével

$$H = \frac{\sum_{i=1}^{n} \left| \frac{x_i}{\sum x_i} - \frac{y_i}{\sum y_i} \right|}{2}$$

A Lorenz-görbe, a Hoover-index és a Gini-együttható összefüggése

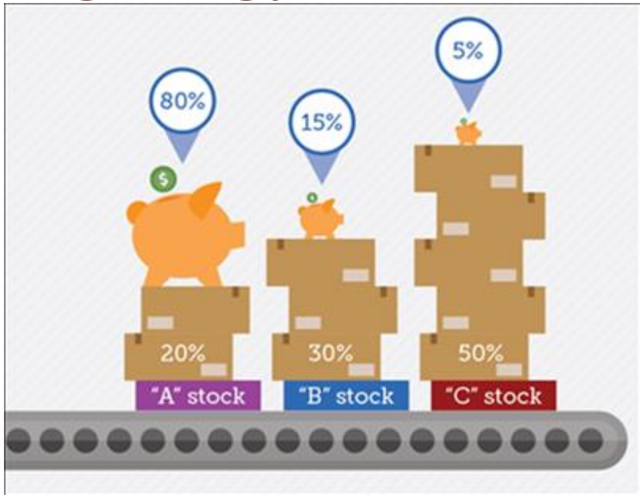
A Lorenz-görbe, a Hoover-index és a Gini-index grafikus interpretációja és a közöttük lévő kapcsolat

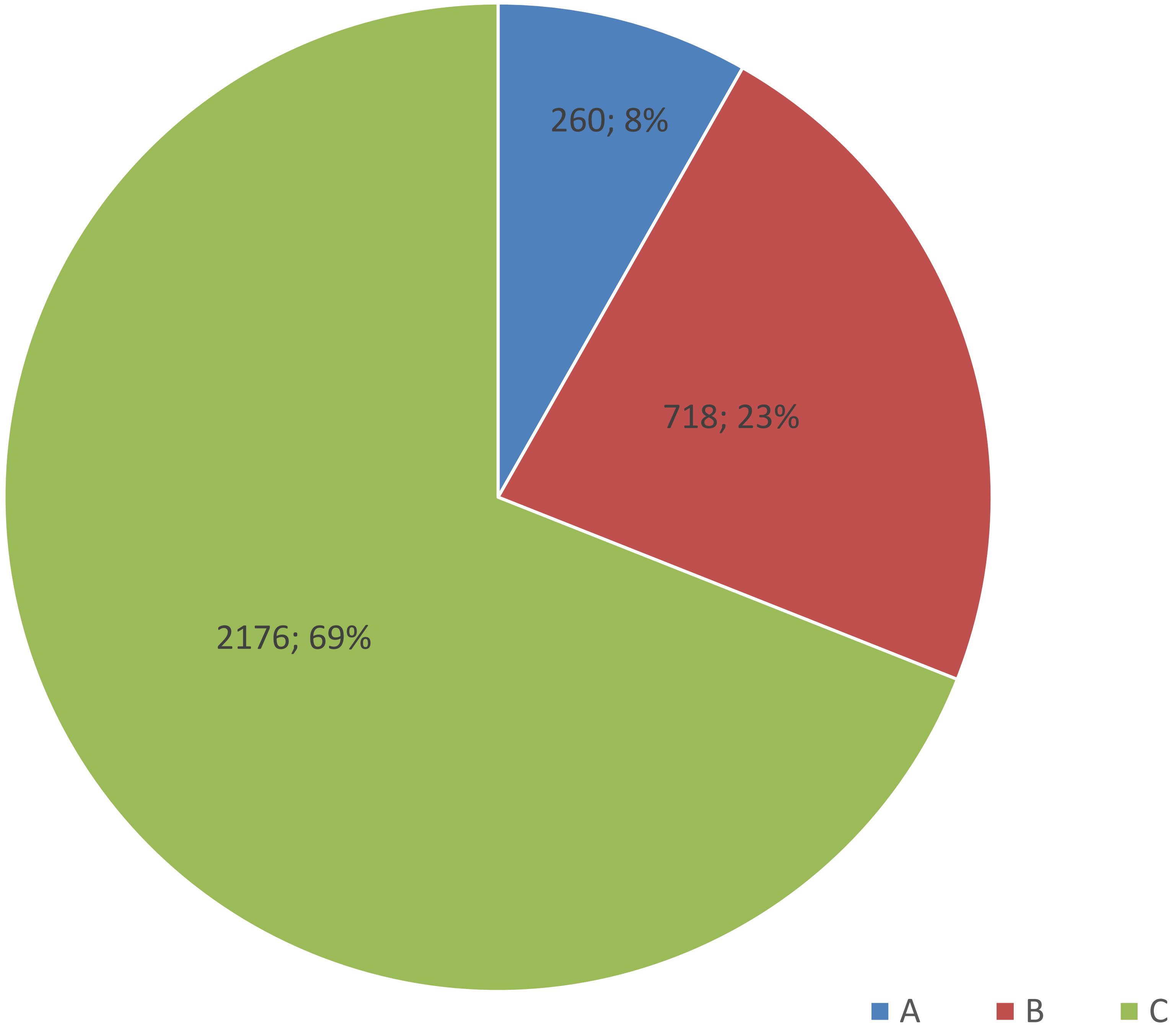


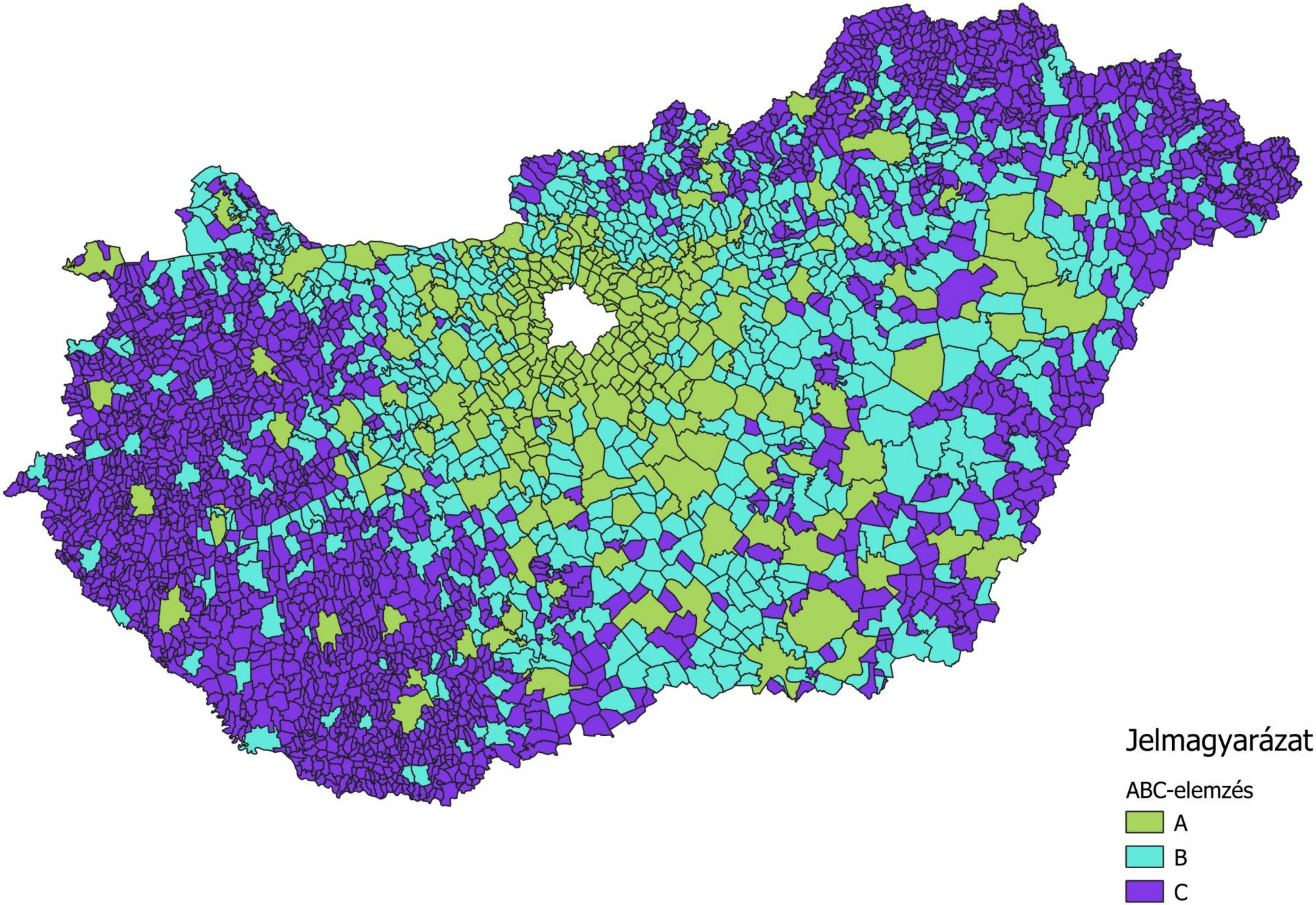
ABC elemzés

- A logisztika területén alkalmazott mutató
- Alapja a Pareto-elv, vagyis hogy a következmények 80%-a az okok 20%-ra vezethetők vissza
- Az egyes „termékek” három csoportba oszthatók:
  - A: Ide tartozik a forgalom 80%-a, míg az egyedek 20%-a
  - B: Ide tartozik a forgalom 15%-a, míg az egyedek 30%-a
  - C: Ide tartozik a forgalom 5%-a, míg az egyedek 50%-a

http://sim.uw.hu/keszletgazdalkodas/index.html?page=16




260; 8%
718; 23%
2176; 69%
A
B
C


Jelmagyarázat
ABC-elemzés
A
B
C

Tartalom

Térstatisztika

↓

Térbeli interakciók

↓

Gravitációs modell

Geográfia első tétele

- Waldo R. Tobler (1930-2018) amerikai-svájci geográfus (1970)
  - Everything is related to everything else, but near things are more related than distant things
  - A világon minden mindennel összefügg, de a közelebbi dolgok jobban hatnak egymásra

Geográfia második tétele

- Waldo R. Tobler (1930-2018) amerikai-svájci geográfus (1999)
  - The phenomenon external to an area of interest affects what goes on inside
  - Egy adott területen belül lévő folyamatokra hatnak a területen kívüli folyamatok

Távolságszámítás

- Egyedek adottak a koordinátáik alapján ($A(A_1; A_2)$, $B(B_1; B_2)$)
- Pitagorasz-tétel megkötései
- Gömbi koszinusz-tétel:

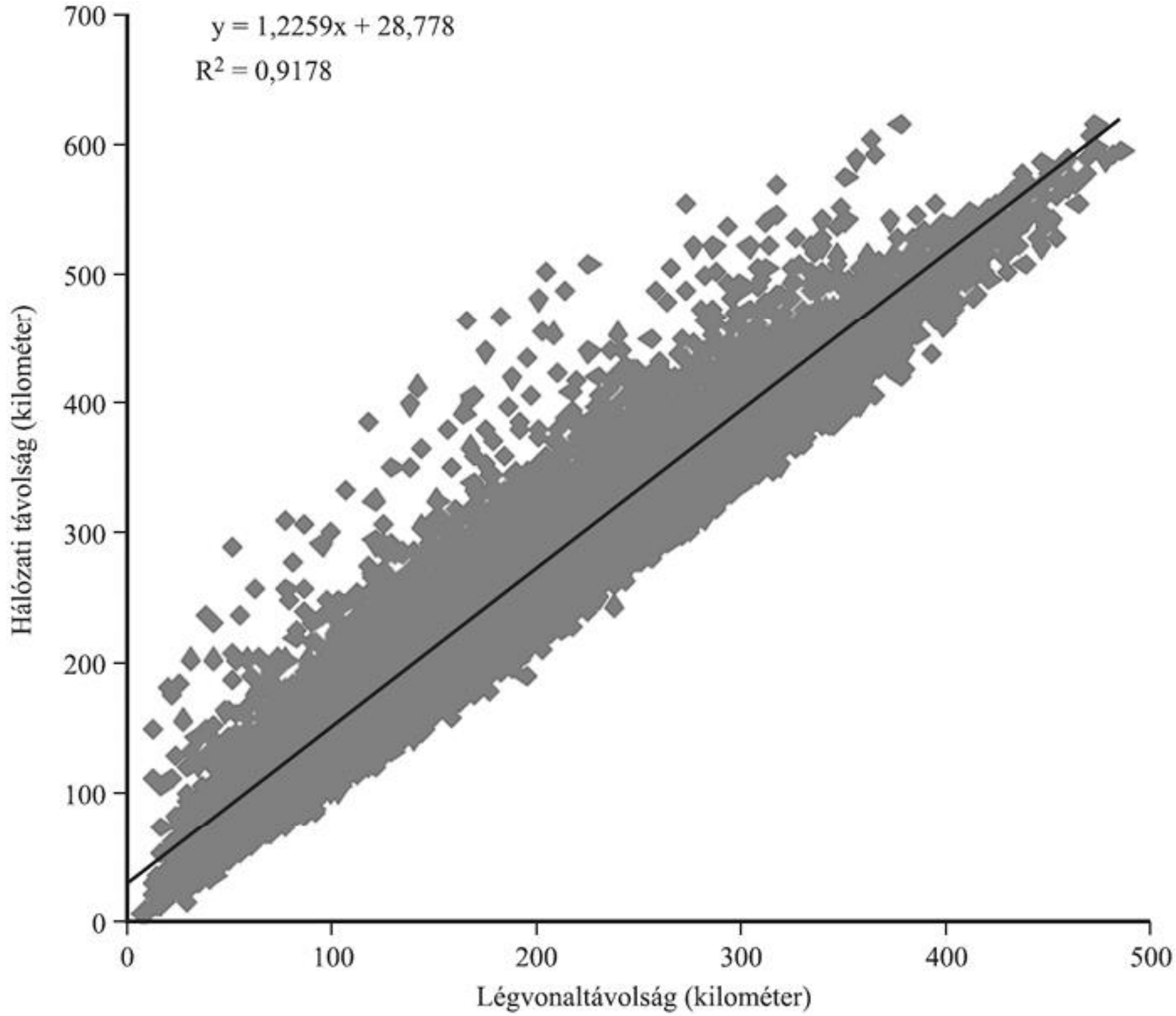
$$\cos c = \cos a \cos b + \sin a \sin b \cos \gamma$$

- Gömbháromszög – két csúcsa a két pont, a harmadik az északi sark

$$a = 90° - B_1$$
$$b = 90° - A_1$$
$$\gamma = |A_2 - B_2|$$

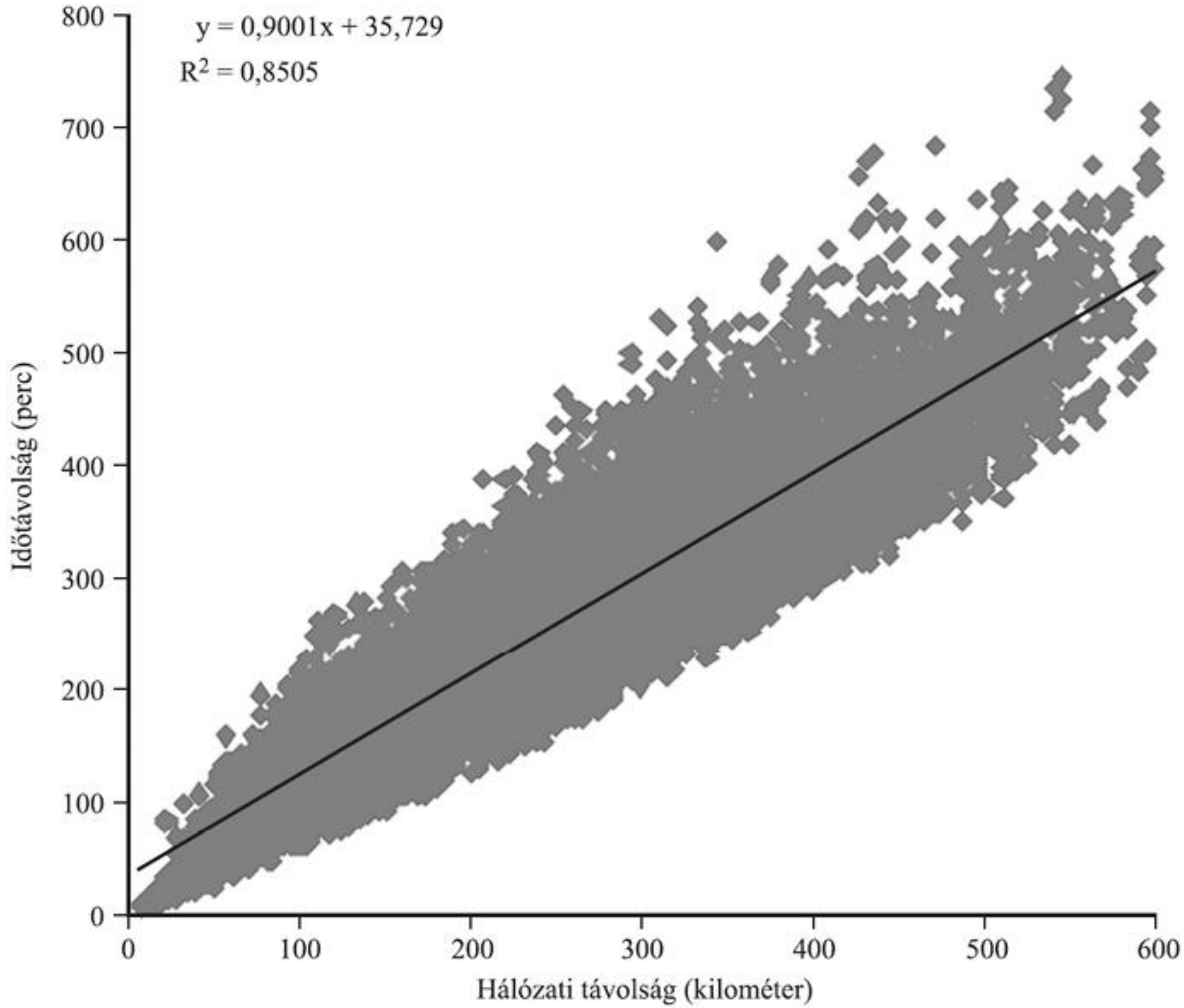
- A két pont távolsága: $d(i, j) = cR$, ahol $R = 6380km$

Hálózati távolság és légvonaltávolság



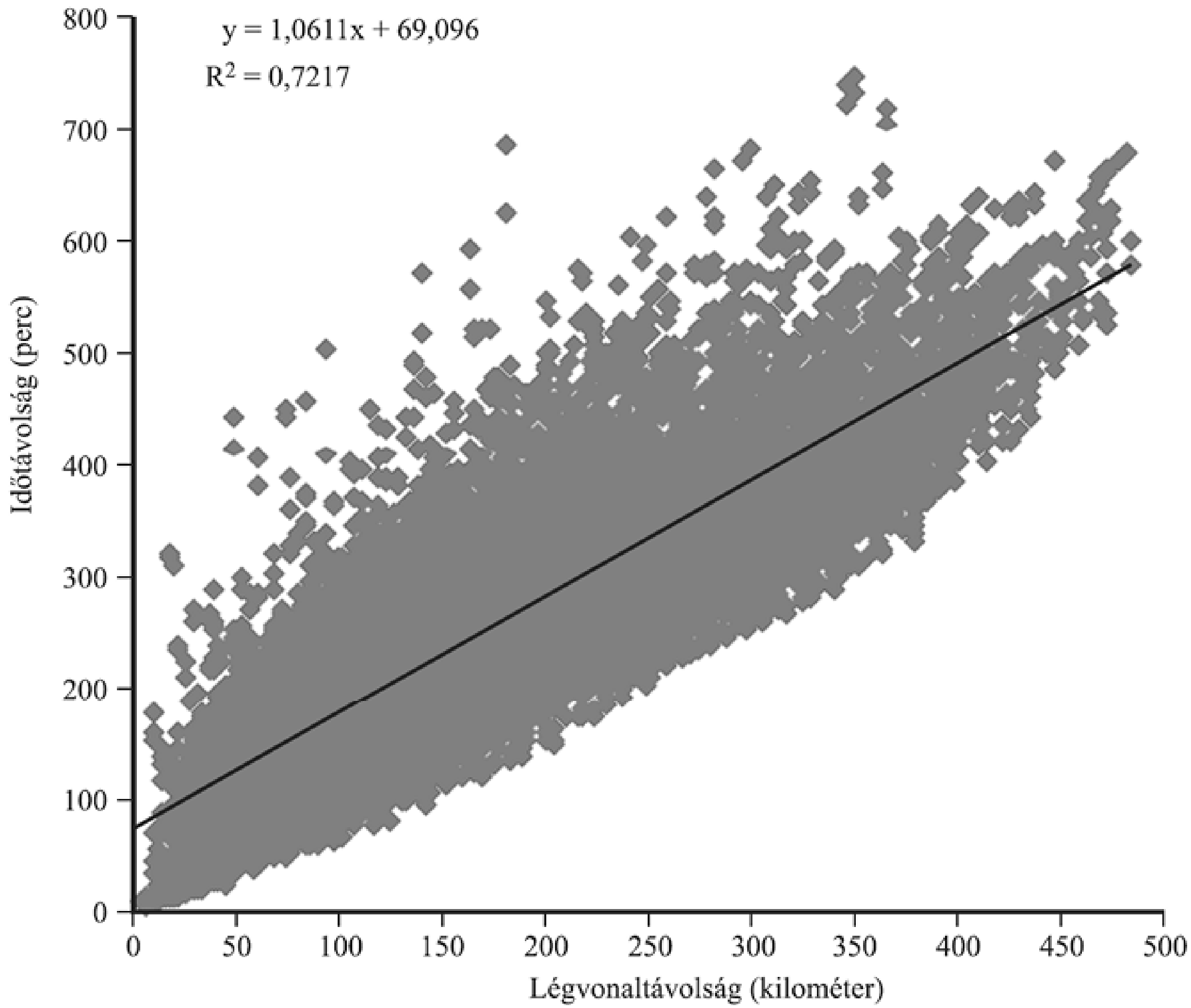
Dusek T., Kotosz B., 2016. Területi statisztika. Akadémiai Kiadó, Budapest.

Időtávolság és hálózati távolság



Dusek T., Kotosz B., 2016. Területi statisztika. Akadémiai Kiadó, Budapest.

Időtávolság és légvonaltávolság



Dusek T., Kotosz B., 2016. Területi statisztika. Akadémiai Kiadó, Budapest.

Legközelebbiszomszéd-index

- Területi index amely a települések mintázatát hasonlítja össze egy elméleti véletlenszerű eloszlással

$$R = \frac{D_{obs}}{D_{exp}}$$

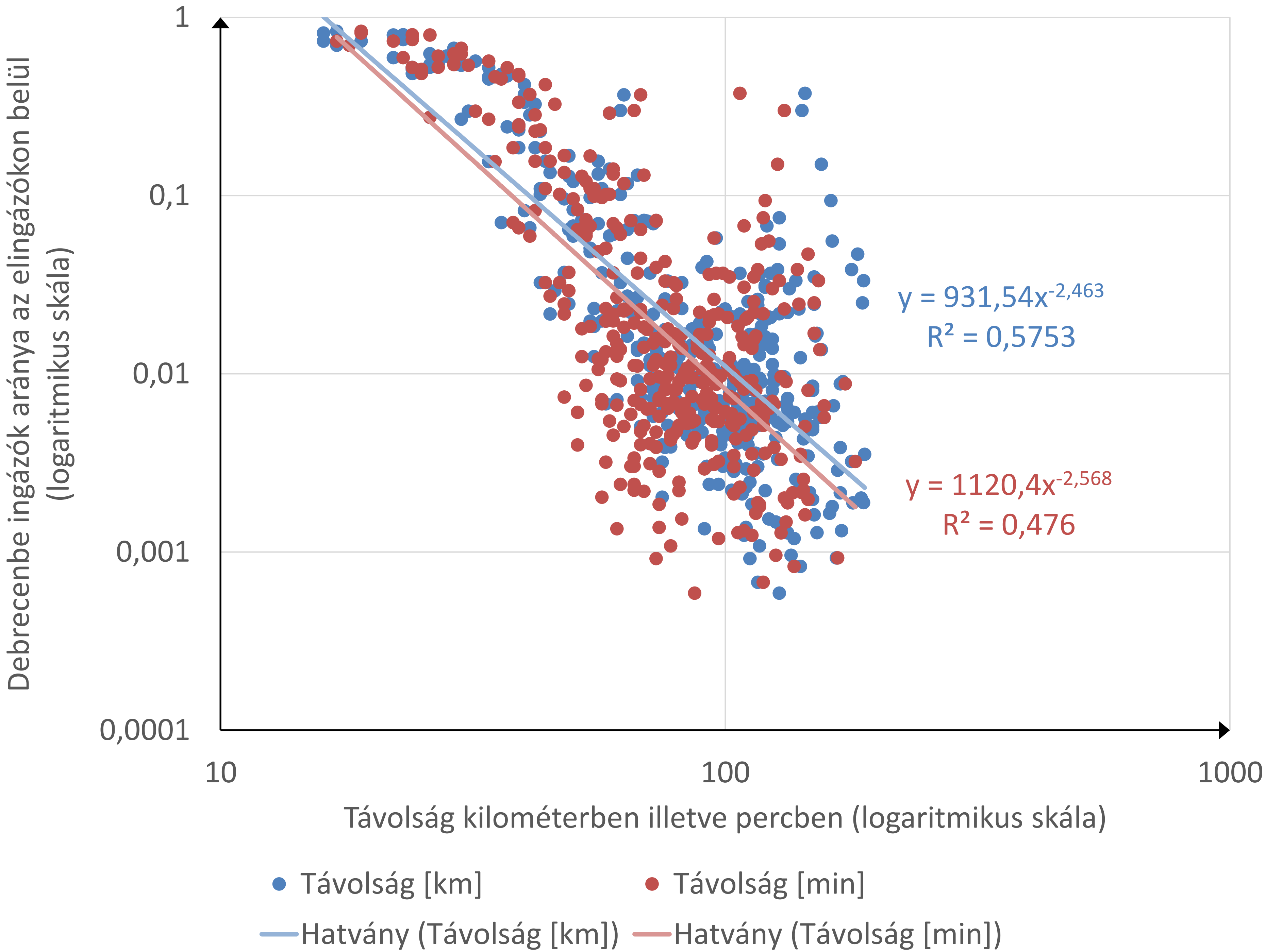
  – ahol:
    - D_{obs}: az összes város, és a hozzá legközelebb lévő város távolságának az átlaga
    - $D_{exp} = \frac{1}{2\sqrt{A}}$ várható átlagos távolság véletlenszerű eloszlás esetén, ahol A az 1 km^2-re jutó városok száma, tehát $[A] = {}^{város}/_{km^2}$
- Ha
  – $R = 0$ akkor minden város egy pontban koncentrálódik
  – $R = 1$ akkor a települések véletlenszerűen oszlanak el
  – $R < 1$ akkor koncentrált elhelyezkedés figyelhető meg
  – $R > 1$ akkor szétszórt vagy szabályos elhelyezkedésről beszélhetünk

Távolság-hanyatlás görbék

- A geográfia első törvényéből adódik, hogy a kölcsönhatás erőssége a távolság növekedésével csökken
- Ezen kapcsolat leírására szolgál a Pareto-függvény

$$F = aD^{-b}$$

  - Ahol
    - F: kölcsönhatás erőssége
    - D: távolság
    - a és b: együtthatók
- A legérdekesebb kérdés b értéke
- Gyakorlatban $b = 2$ alkalmazandó leggyakrabban
- Svéd kutatók migráció elemzése során $b \in [0{,}4..3{,}3]$ értéket tapasztaltak 1,94 középértékkel
- Példa: Debrecenbe naponta ingázók száma az egyes településekből Hajdú-Bihar, valamint a szomszédos megyékből



1
0,1
0,01
0,001
0,0001
10
100
1000
Debrecenbe ingázók aránya az elingázókon belül (logaritmikus skála)
Távolság kilométerben illetve percben (logaritmikus skála)
y = 931,54x-2,463
R² = 0,5753
y = 1120,4x-2,568
R² = 0,476
Távolság [km]
Távolság [min]
Hatvány (Távolság [km])
Hatvány (Távolság [min])

Tartalom

Térstatisztika

Térbeli interakciók

Gravitációs modell

Bevezetés

- A gravitációs modellek esetében a távolság-hanyatlás görbéknél megismert α paramétert vizsgáljuk tovább
- Ehhez bevonjuk a vizsgálatba a két végpontot, vagyis a kiinduló, valamint a célállomást
- Az elemzés alapja pedig egy fizikai törvényszerűség

Gravitációs modellek

- A távolság-hanyatlás görbéknél láthattuk, hogy a kapcsolat kitevőjének középértéke 2 körül alakul (1,94)
- Ez azért érdekes, mert összhangban van a Newton-féle gravitációs összefüggéssel

$$\boldsymbol{F} = -\gamma \frac{m_1 m_2}{r^3} \boldsymbol{r}$$

  - $\boldsymbol{F}$: tömegvonzási erő
  - m_1, m_2: a két test tömege
  - r: a két test közti távolság ($\boldsymbol{r}$ értelemszerűen a távolságvektor, amely ellentétes irányú a tömegvonzási erő vektorával)
  - γ: Newton-féle gravitációs együttható ($\gamma = 6{,}67259 * 10^{-11} \frac{Nm^2}{kg^2}$)

Gravitációs modellek

- Gyakorlatban, ha nem vektorokkal, csak a nagyságukkal számolunk, akkor
$$F = \gamma \frac{m_1 m_2}{r^2} = \gamma m_1 m_2 r^{-2}$$
- Ezt általánosíthatjuk
  - Valamilyen kölcsönhatást, vagy térbeli igényt feltételezünk (legyen U) (például utazási igény)
  - Valamilyen térbeli egységre (i) jellemző súlyt definiálunk (m_i) (például lakosságszám, GDP)
  - Valamilyen távolságot definiálunk (r) (például légvonalbeli, közúti, eljutási idő)
  - Adódik egy együttható (legyen γ)

Gravitációs modellek

- Háromféle megközelítés alkalmazható
  - Csak az együttható értékét becsüljük, a többit adottnak tekintjük
    - Ekkor $y = bx$ konstans tag nélküli lineáris függvényt kell becsülni
  - Az együtthatót és a távolság kitevőjét becsüljük
    - Ebben az esetben a kölcsönhatás iránya nem számít

$$U = \gamma(m_1 m_2)^{\beta_1} r^{\beta_2} \rightarrow \ln U = ln\gamma + \beta_1 \ln(m_1 m_2) + \beta_2 \ln r$$

    - A nyíl utáni egyenlet lineáris regresszióként becsülhető
  - Mindent becslünk
    - Ebben az esetben a kölcsönhatás iránya számít
    - Logaritmizálás után a következő egyenlet adódik, mely lineáris regresszióként becsülhető

$$\ln U = \ln\gamma + \beta_1 \ln m_1 + \beta_2 \ln m_2 + \beta_3 \ln r$$

Egyéb tényezők figyelembevétele

- A klasszikus felírásból kiindulva megfigyelhető, hogy a modell alakja zárt, nehéz más tényezők figyelembevétele
- Azonban az együttható alkalmazása lehetővé tesz egyéb magyarázóváltozók (x_4, x_5 stb.) figyelembevételét
- ! $U = \gamma m_1^{\beta_1} m_2^{\beta_2} r^{\beta_3}$
- ! x_4 és x_5 két magyarázóváltozó amely döntően befolyásolja U értékét

$$\ln U = \alpha + \beta_1 \ln m_1 + \beta_2 \ln m_2 + \beta_3 \ln r + \beta_4 x_4 + \beta_5 x_5$$

- Ebben az esetben

$$\gamma = f(\alpha, x_4, x_5) = e^{\alpha} x_4^{\beta_4} x_5^{\beta_5}$$

Gyakorlati példa – Velencei-tó

- Projekt – Velencei-tó környezetében megfigyelhető napi ingázás
- Népszámlálási adatokból tudjuk a járásközpontokba (Bicske, Szfv., Gárdony, Martonvásár) valamint Budapestre ingázók számát
- Ebből teljes OD mátrixot becslünk kétféle becslőmodellel

Modellek

$$U = e^{\alpha}(m_1 m_2)^{\beta_1} r^{\beta_2} c^{\beta_3} h^{\beta_4} j^{\beta_5}$$
$$\ln U = \alpha + \beta_1 \ln(m_1 m_2) + \beta_2 \ln r + \beta_3 \ln c + \beta_4 \ln h + \beta_5 \ln j$$

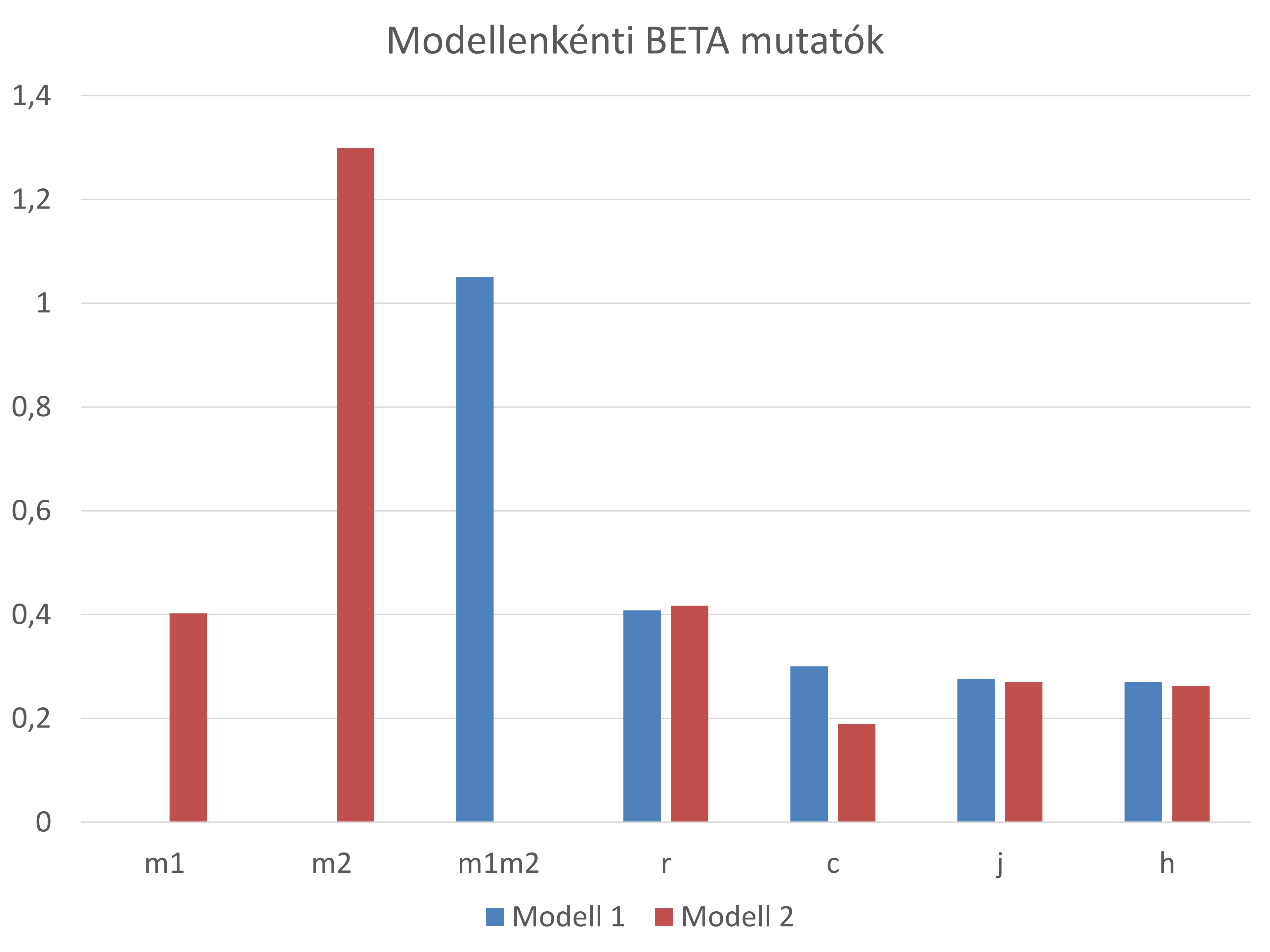
$$U = e^{\alpha} {m_1}^{\beta_1} {m_2}^{\beta_2} r^{\beta_3} c^{\beta_4} h^{\beta_5} j^{\beta_6}$$
$$\ln U = \alpha + \beta_1 \ln m_1 + \beta_2 \ln m_2 + \beta_3 \ln r + \beta_4 \ln c + \beta_5 \ln h + \beta_6 \ln j$$

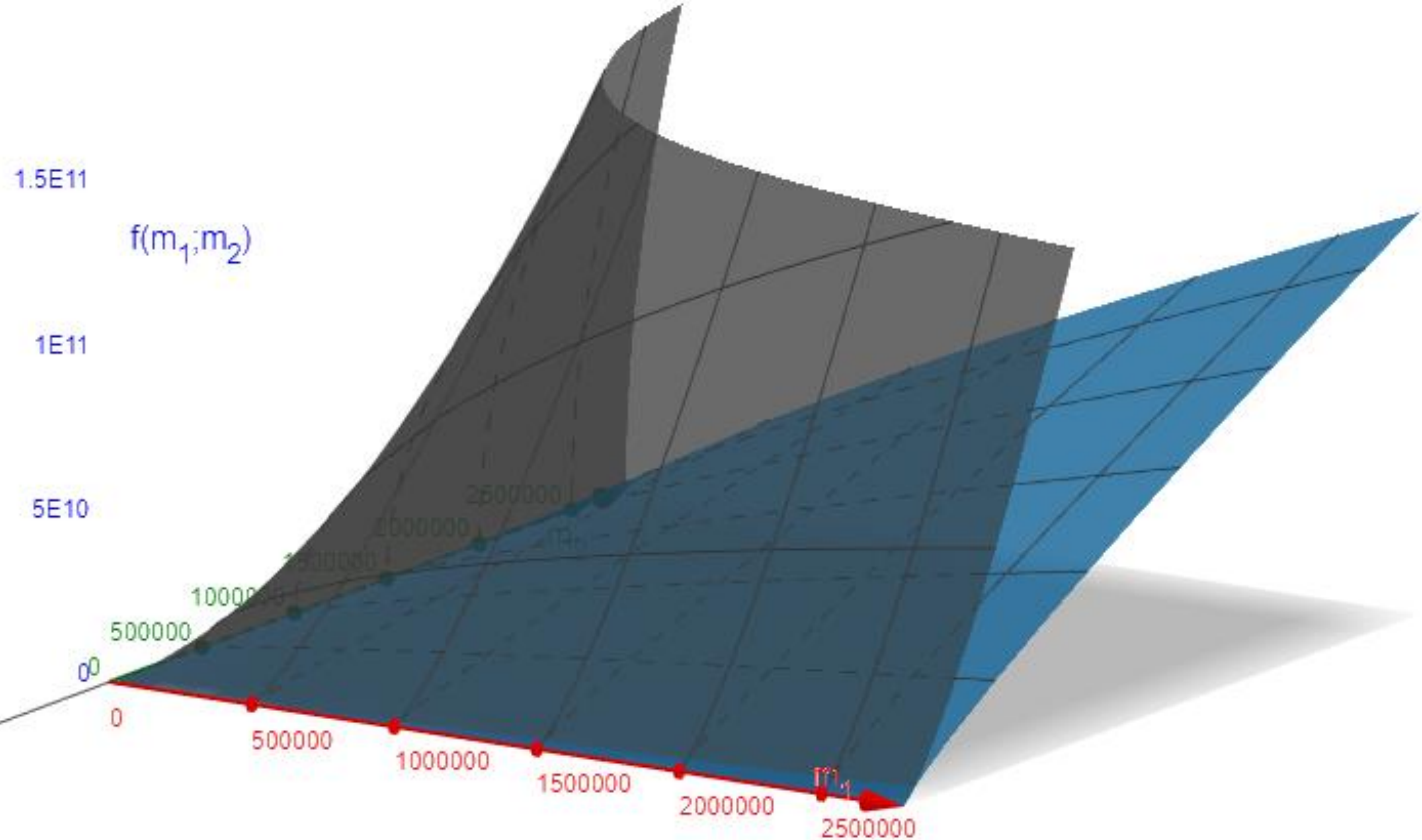
- Ahol:
  - U: ingázók száma
  - m_1: kiinduló település lakosságszáma
  - m_2: érkező település lakosságszáma
  - c: a kiinduló és az érkező település rangjának hányadosa
  - h: érintett megyék száma
  - j: érintett járások száma

Paraméterbecslés

Tengelymetszet	-3,9847	
	[-2,4811]	*
$\ln(m_1 m_2)$	0,8598	
	[12,8762]	***
$\ln r$	-1,9787	
	[-8,0645]	***
$\ln c$	0,9401	
	[5,6546]	***
$\ln j$	-1,3894	
	[-3,9133]	***
$\ln h$	-1,9247	
	[-3,0264]	**
R-négyzet	0,8079	
Standard hiba	0,9449	
F-próba	[91,7085]	***

Tengelymetszet	-5,1249	
	[-3,2970]	*
$\ln m_1$	0,5372	
	[4,9362]	***
$\ln m_2$	1,2872	
	[9,6573]	***
$\ln r$	-2,0220	
	[-8,6816]	***
$\ln c$	-0,5910	
	[-1,3170]	
$\ln j$	-1,3588	
	[-4,0356]	***
$\ln h$	-1,8762	
	[-3,1112]	**
R-négyzet	0,8290	
Standard hiba	0,8958	
F-próba	[87,2461]	***


Modellenkénti BETA mutatók
1,4
1,2
1
0,8
0,6
0,4
0,2
0
m1
m2
m1m2
r
c
j
h
Modell 1
Modell 2


1.5E11
f(m1;m2)
1E11
5E10
0
0
500000
1000000
1500000
2000000
2500000
0
500000
1000000
1500000
2000000
2500000
m1

Ellenőrző kérdések

- Milyen statisztikai szempontú fenntartásaink lehetnek a téradatokkal kapcsolatosan?
- Milyen következményei vannak az adatok nem normális eloszlásának?
- Mit mond ki a geográfia első tétele?
- Mi a gravitációs modell? Hogy lehet az együtthatókat becsülni, és mi ennek az oka?

KÖSZÖNÖM A MEGTISZTELŐ FIGYELMÜKET!

Dr. Sipos Tibor
Dr. Török Ádám
Szabó Zsombor

MŰEGYETEM 1782

KUKG