

1. alkalom

Adminisztratív információk, alapfogalmak

Budapesti Corvinus Egyetem



Budapesti Corvinus Egyetem
Corvinus University of Budapest

Tartalom



Adminisztratív információk

Sokaság

Ismérv

Mérési skálák

Adatfelvétel

Statisztikai alapműveletek



Adminisztratív információk



Oktatók, elérhetőség információforrások



- Oktató: Tőkés László (Közgazdaságtan Intézet)
- Fogadóóra: szerda, 13:40-15:10, E.219.
- Email: laszlo.tokes@uni-corvinus.hu
- Információk:
 - Moodle
 - Tantárgyi adatlap és tematika

Tananyag



- Tankönyv: G. Koop (2008): Közgazdasági adatok elemzése (Osiris Kiadó)
- Az oktatók által kiadott egyéb cikkek és feladatok
- Szoftver: Excel

Számonkérés



Követelmény:

10 db „1 perces” dolgozat: 10x0,5 pont

10 db egyéni órai feladat beadása: 10x0,5 pont

10 db egyéni házi feladat: 10x1 pont

2 db csoportos házi feladat a félév során: 2x5 pont

2 db negyedéves dolgozat: 2x35 pont (minimumköv.: 50%-50%)

Összesen: 100 pont

- Aláírás feltétele: 7 darab egyéni házi feladat beadása határidőre.

Adatelemzés, ökonometria



1. Adatgyűjtés
 2. Adatok jellemzése
 3. Összefüggések feltárása
- Statisztika: adatok elemzésével foglalkozó tudomány.
 - Ökonometria: közgazdasági modellek alapján adatok elemzésével foglalkozó tudomány. („közgazdasági” itt tágan értelmezendő – pl. egészség-, oktatáspolitikai és társadalomtudományok)

Tantárgy célja: adatelemzés alapjai + kedvet csinálni

Mit csinál és mit nem a kurzus?



- Nem lesz statisztikai elmélet és ökonometria. (Ezek későbbi kurzusok anyaga).
- Nem szakácskönyvet tanulunk.
- Korrelációt és regressziót járjuk alaposan körbe, majd idősoelemzéssel is foglalkozunk kicsit.
- Adatok adottak. Nem foglalkozunk azzal, hogy hogyan is állnak elő.
- Sok példa lesz.

Szoftver



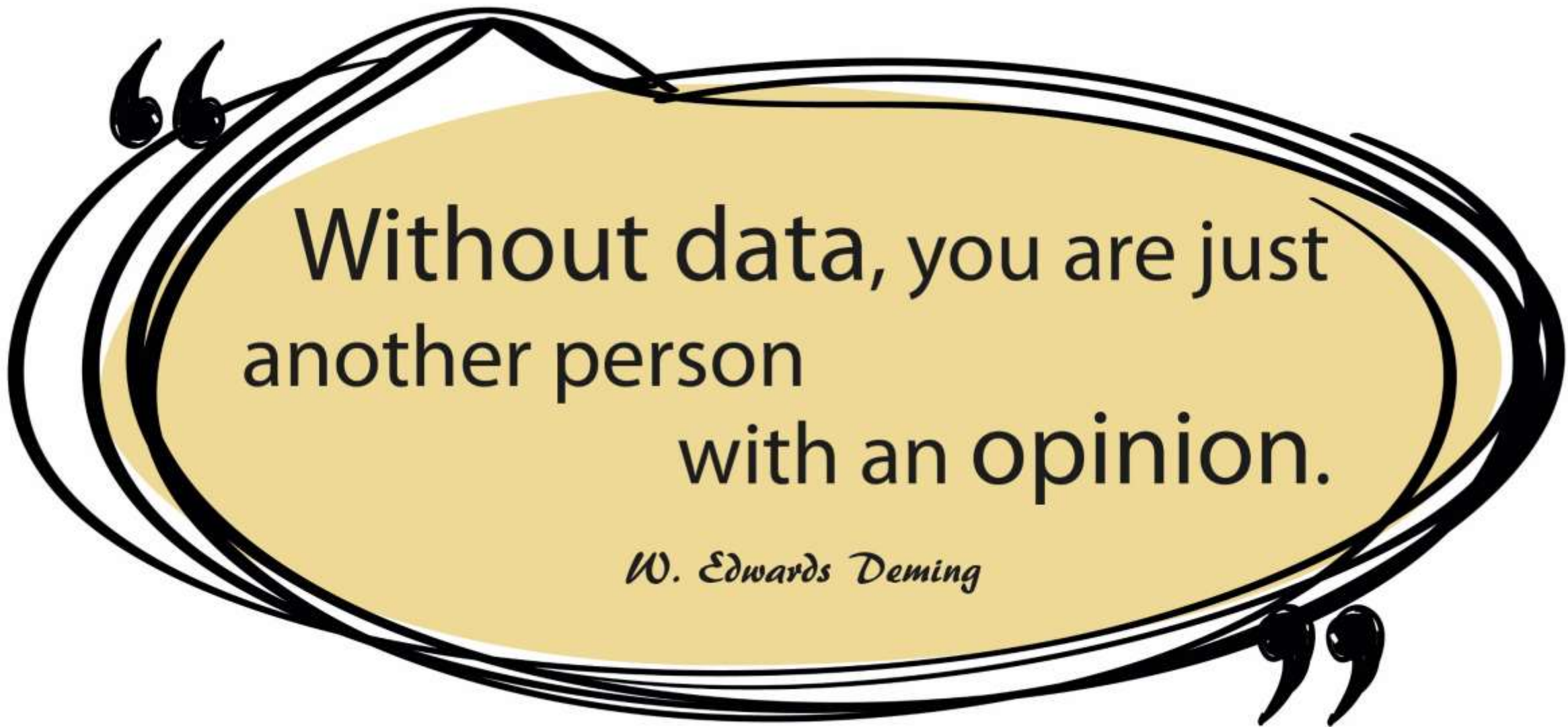
- Excelt mindenkinek nagyon kell tudnia, mert az mindenhol alap.
- Excelen túl jó néhány szoftvert alaposan ismerni (Stata, R, Matlab, Gretl, Eviews, stb.).
- Érdemes más kurzusokon is energiát fektetni a programok megismerésébe
Itt mi most csak Excelt használunk.



Bevezetés

alkalom



A hand-drawn speech bubble with a yellow fill and a thick black outline. The bubble is irregular in shape, with a pointed top and a rounded bottom. It contains a quote in a sans-serif font. The quote is centered and reads: "Without data, you are just another person with an opinion." Below the quote, the name "W. Edwards Deming" is written in a smaller, italicized font. The bubble is decorated with large, stylized black quotation marks at the top left and bottom right corners.

Without data, you are just
another person
with an opinion.

W. Edwards Deming



- Megfigyelés vagy tapasztalatok alapján kialakul bennünk egy kép, hogy **hogyan működik a világ egy minket érdeklő szelete.**
- Ha ezt a képet jól átgondoljuk, élünk bizonyos feltevésekkel (a világ sokkal bonyolultabb, mintsem hogy mindent figyelembe vegyünk) és azok alapján leszűrünk bizonyos **következtetéseket**, akkor lesz egy **közgazdasági elméletünk.**
- A jó elmélet egyik ismérve az, hogy adatokon ellenőrizhető, azaz **tesztelhető** előrejelzéseket ad.
- Rengeteg lehetséges elmélet van mindenre, az **adatokkal való összevetés dönti el, hogy melyik a legjobb.**



- Miután megvan a kidolgozott elméletünk, **adatokat gyűjtünk**. Sokszor nincs pont olyan adat, amire szükségünk van, akkor olyanhoz kell folyamodni, ami részben pótolja, kifejezi a hiányzó adatot.
- Majd a meglévő **adatokon teszteljük az elmélet helyességét**. Illetve sokszor azt teszteljük, hogy a versengő elméletek közül melyik írja le jobban a valóságot.
- **Elmélet és adatelemzés adja a teljes közgazdasági elemzést**. Adatelemzés nélkül az elmélet kissé üres, hiszen általában könnyen elképzelhetőek alternatív elméletek. Elmélet nélküli adatelemzés (pl. adatbányászat – data mining) azért nem az igazi, mert az a jó, ha van egy elképzelésünk, hogy két dolog miért mozog együtt, hogyan hat az egyik a másikra.



- Az adatok elemzése során a legtöbbször nem azt találjuk, hogy az adatok tökéletesen alátámasztják az elméletet. Ha ilyet látunk, akkor vagy valami teljesen triviális dologra bukkantunk, vagy valamit elrontottunk. A világ sokkal bonyolultabb annál, hogy tökéletesen meg tudjuk magyarázni.

Adatok szerepe a közgazdaságtudományban



TABLE 4
PERCENT DISTRIBUTIONS OF METHODOLOGY OF PUBLISHED ARTICLES, 1963–2011*

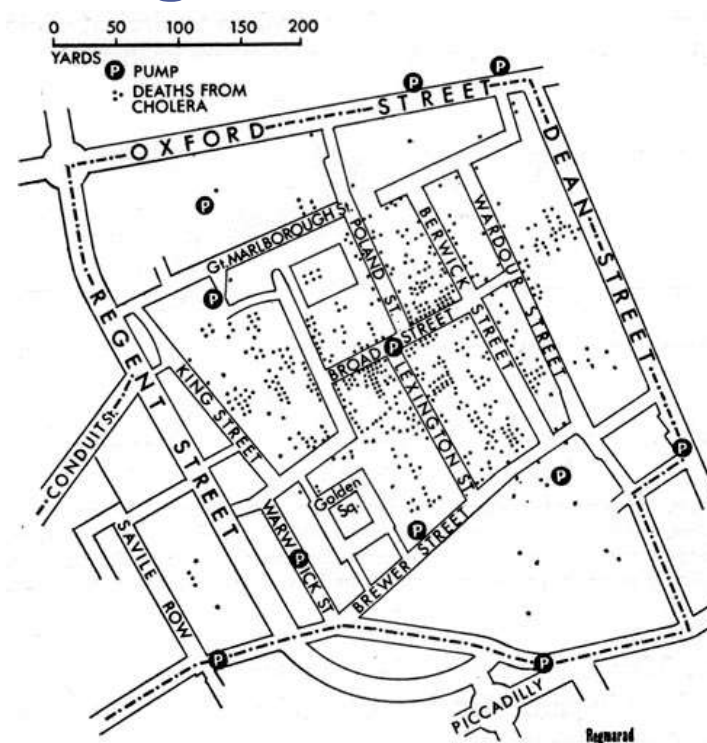
Year	Type of study				
	Theory	Theory with simulation	Empirical: borrowed data	Empirical: own data	Experiment
1963	50.7	1.5	39.1	8.7	0
1973	54.6	4.2	37.0	4.2	0
1983	57.6	4.0	35.2	2.4	0.8
1993	32.4	7.3	47.8	8.8	3.7
2003	28.9	11.1	38.5	17.8	3.7
2011	19.1	8.8	29.9	34.0	8.2

Forrás: Hamermesh, DS (2013): Six Decades of Top Economics Publishing: Who and How? Journal of Economic Literature 51 (1): 162-172.

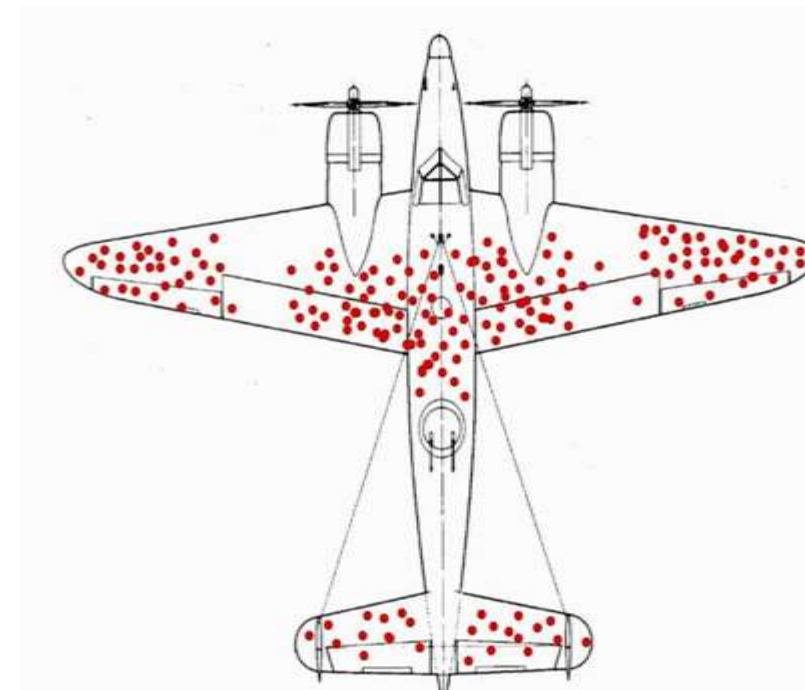
Adatelemzés: Az elefántcsonttoronybéli tudósok haszontalan, úri mulatsága?



Sumer birodalom:
népszámlálás az ételkészletek
hatékony elosztása érdekében.



John Snow: kolerajárvány
megfékezése Londonban.



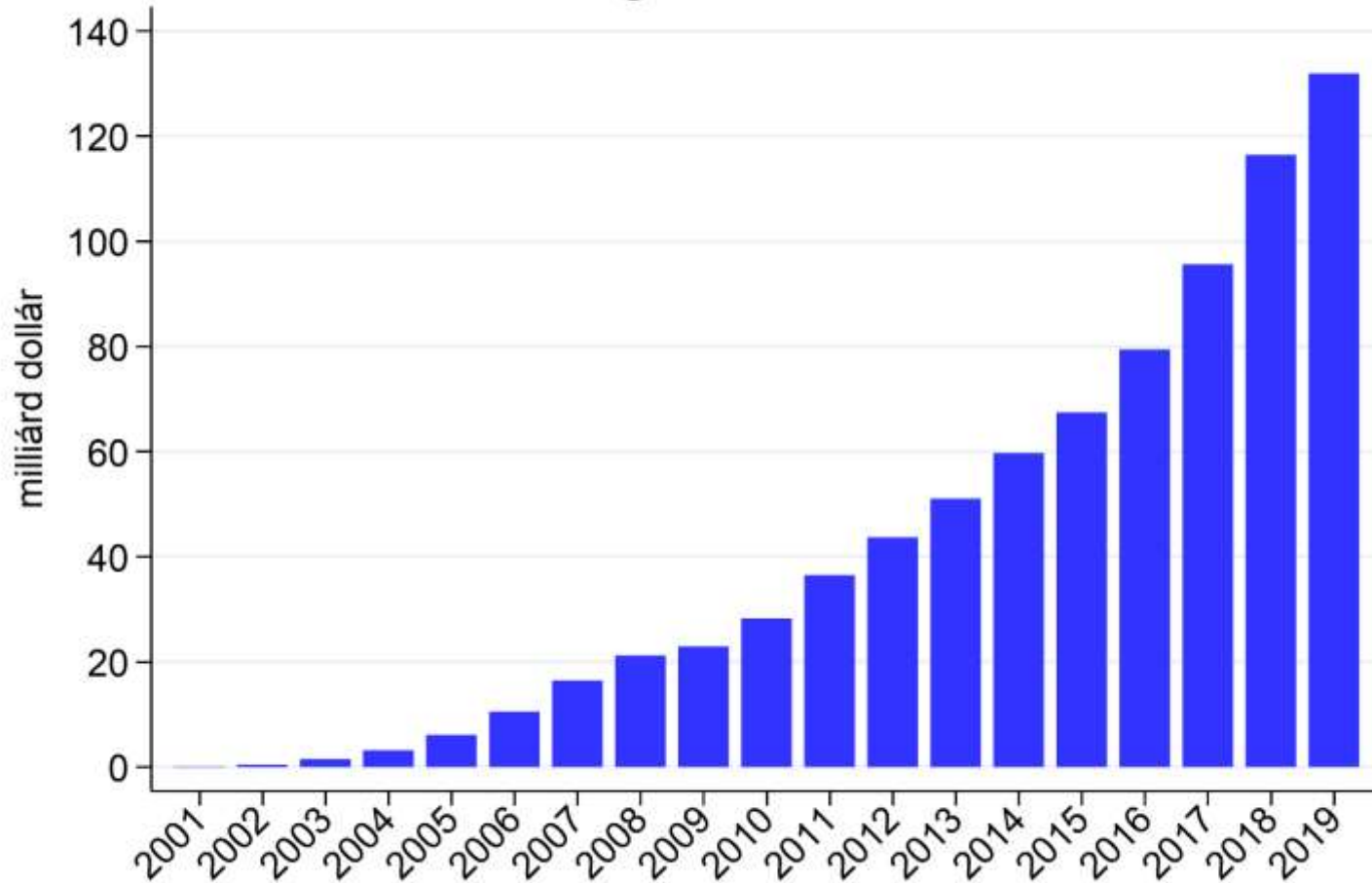
Wald Ábrahám: harci repülők
páncélzatának fejlesztése a 2.
világháborúban.

Forrás: <https://processgold.com/blog/top-10-examples-of-successful-data-analysis/>

Adatelemzés: Az elefántcsonttoronybéli tudósok haszontalan, úri mulatsága?



A Google reklámbevétele



Forrás: Statista. Megjegyzés: Folyóáras adatok.

- „Search ads use your activity to be more personally relevant.”
- „Gmail ads are based on your activity.”
- „Google Play ads help you discover apps you might like.”
- „YouTube ads use your search and watch activity to be more relevant.”

Magyarország bruttó hazai terméke (GDP) 2018-ban:

157,8 milliárd dollár

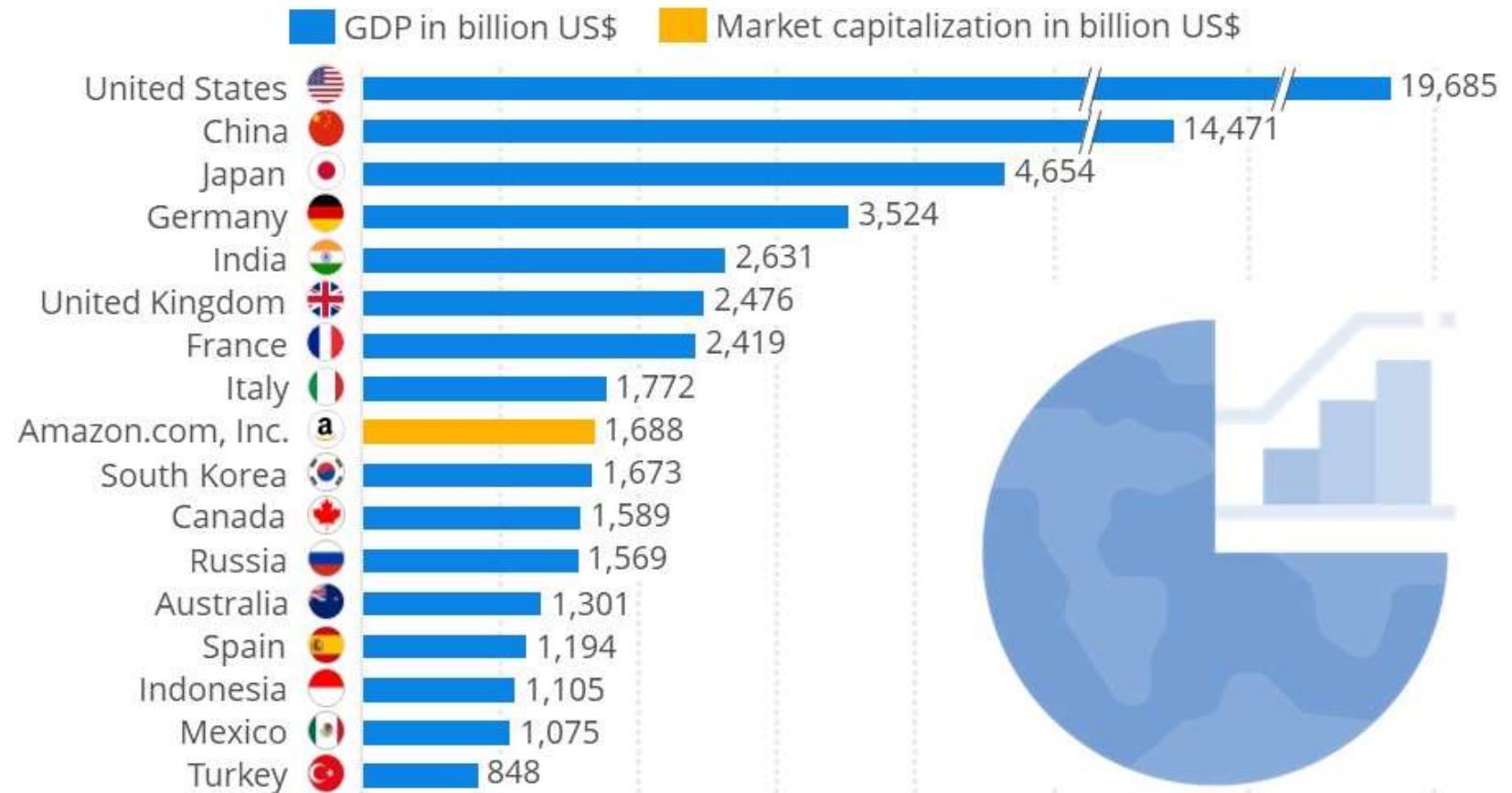
Adatelemzés: Az elefántcsonttoronybéli tudósok haszontalan, úri mulatsága?

Forrás: <https://ecommercedb.com/en/blogPost/3256/amazon-market-cap-beats-gdp-of-south-korea>



Amazon's market cap is higher than the GDP of whole countries

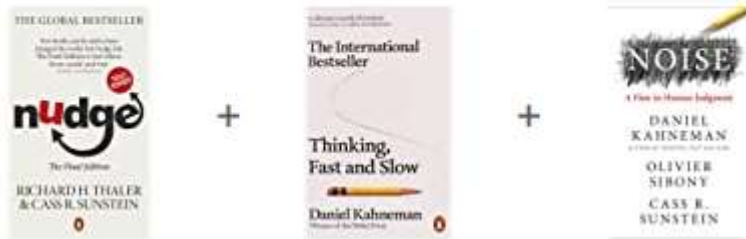
GDP¹ in selected countries and market capitalization of Amazon.com, Inc.²



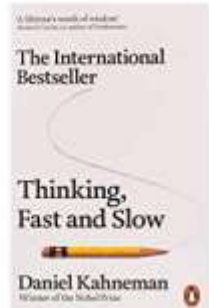
1: Gross Domestic Product (GDP) in billion US\$ in 2020 2: Market capitalization of Amazon.com, Inc. in billion US\$ as of 08/02/2021
Sources: Statista Country Outlook, Yahoo Finance



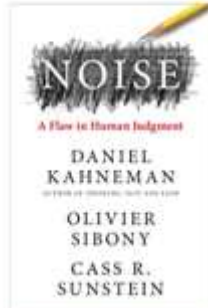
Frequently bought together



Customers who viewed this item also viewed



Thinking, Fast and Slow: Daniel Kahneman
» Daniel Kahneman
★★★★☆ 20,300
Paperback
£9.99



Noise: The new book from the authors of...
» Daniel Kahneman
★★★★☆ 505
Hardcover
£16.55



Alapfogalmak



Alapfogalmak: statisztika

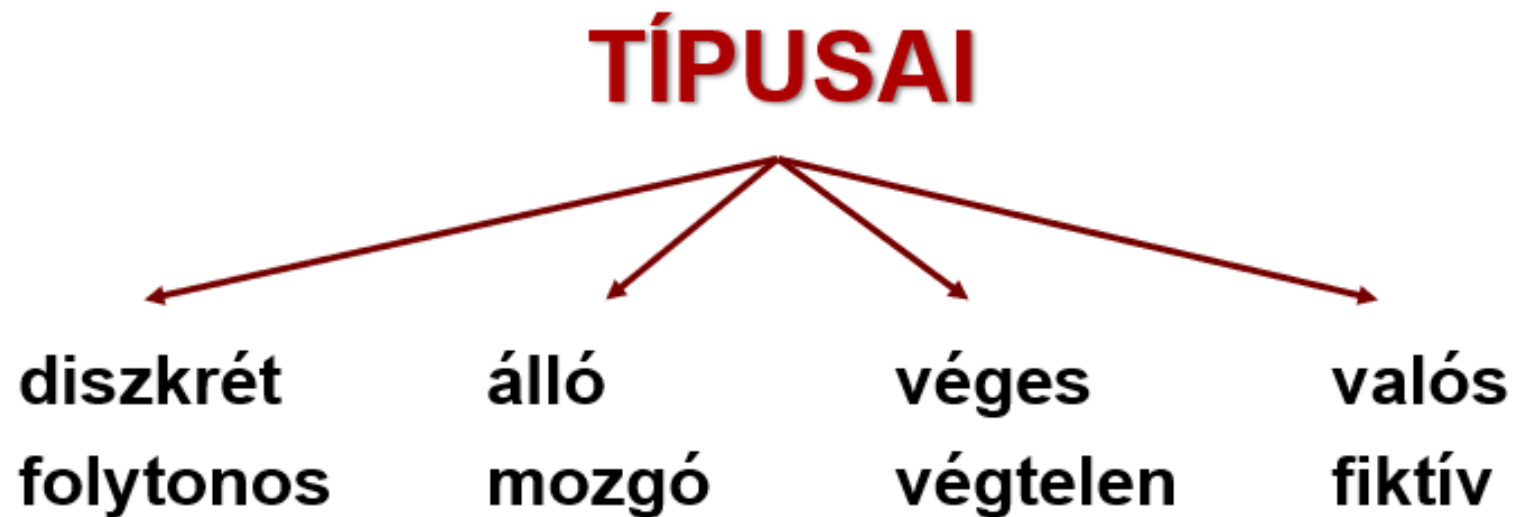


- A valóság tényeit tömören, számszerűen jellemző, modellezni törekvő:
 - gyakorlati (adatgyűjtő, elemző) tevékenység,
 - tudományos módszertan
- Tömegesen előforduló jelenségek tömör, számszerű jellemzése:
 - adatokkal
 - mutatószámokkal

Alapfogalmak: sokaság (populáció)



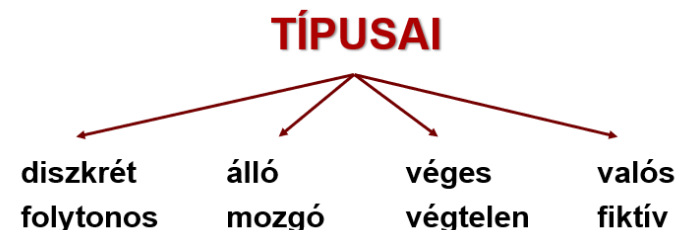
- A vizsgálat tárgyát képező egységek összessége, halmaza.



Alapfogalmak: sokaság (populáció)



- Diszkrét vs. folytonos:
 - *Diszkrét*: jól elkülönülő egységekből áll.
 - *Folytonos*: csak önkényesen elkülöníthető egységekből áll.
- Álló vs. mozgó:
 - *Álló*: időpontra vonatkozó (állományjellegű, stock).
 - *Mozgó*: időtartamra vonatkozó (folyamatjellegű, flow).
- Véges vs. végtelen: számosságra utal.
- Valós vs. fiktív:
 - *Valós*: létező, ténylegesen előforduló egységekből áll.
 - *Fiktív*: elképzelt egységekből áll.

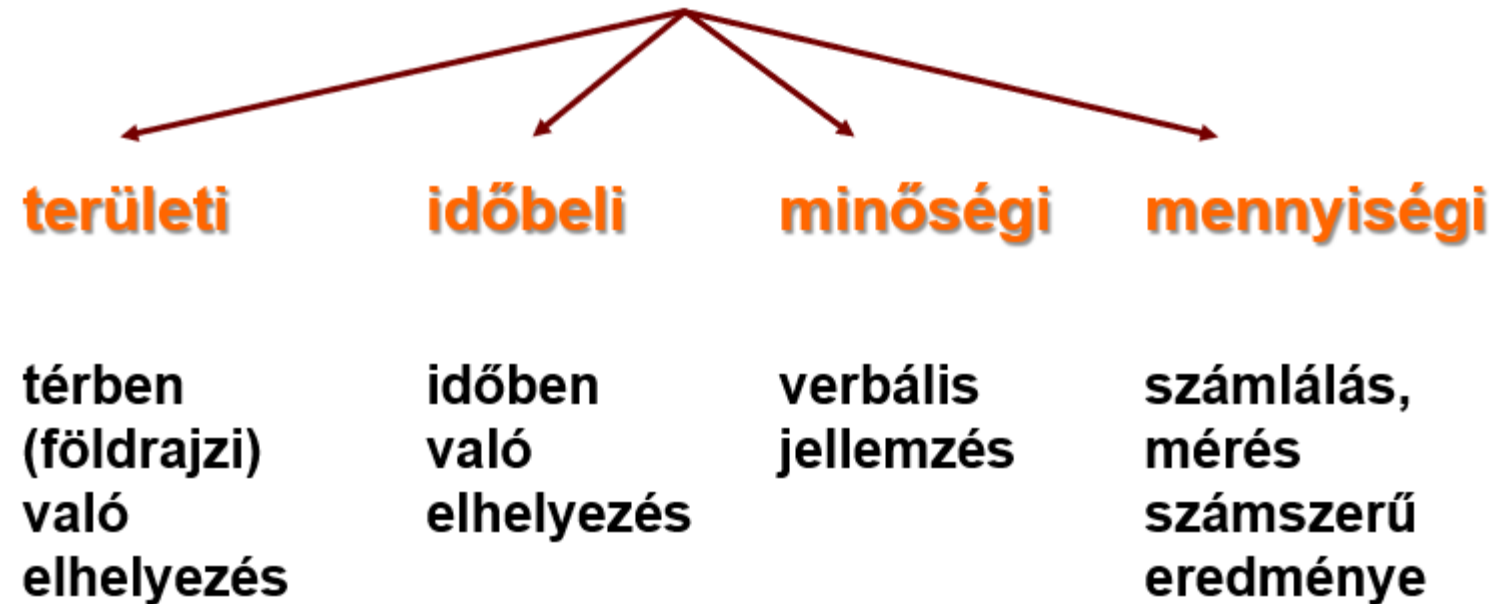


Alapfogalmak: ismérv



- *Ismérv*: Olyan vizsgálati szempont, amely alapján a sokaság részekre bontható.
- A valamely adott szempont szerint lehetséges tulajdonságokat *ismérvváltozatoknak* nevezzük.

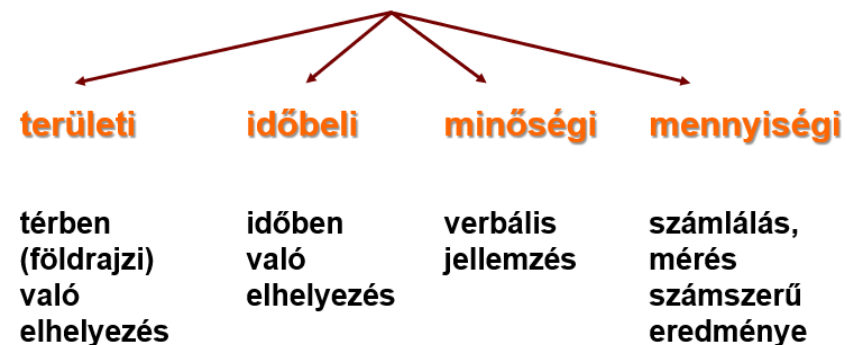
Az ismérvváltozatok által adott információ természete alapján négyféle ismérvfajtát különböztetünk meg:





Alapfogalmak: ismérvek fajtái

- Területi: Az egységek térbeli (földrajzi) elhelyezkedésére utal.
- Időbeli: Az egységek időbeli elhelyezkedésére utal.
- Minőségi: A sokaság egységeit verbálisan jellemzik.
- Mennyiségi: A sokaság egységeihez valamilyen mérés vagy számlálás eredményét rendeli hozzá.



Minőségi vs. mennyiségi adatok



- Kvantitatív: mennyiségi → szám
 - pl. infláció, jövedelem, munkanélküliség
- Kvalitatív: minőségi → nem szám – számokká alakítjuk
 - Férfi/nő - kétértékű változó (dummy): 0-1
 - Iskolai végzettség: alapfok – 1, középfok – 2, felsőfok – 3
 - Figyelni kell, hogy számoknak milyen a viszonyuk, pl. a középfok 2 száma nem azt jelenti, hogy az az alapfok kétszerese.

Alapfogalmak: mérési skálák



- Mérés: sokasági egységek tulajdonságainak szám formájában történő rögzítése.



Alapfogalmak: mérési skálák



- Névleges (nominális): az egységekhez rendelt számérték *egyező* vagy *különböző* voltát engedi meg az egységek ténylegesen is jellemző tulajdonságaként elfogadni. (pl. név, nem, adószám)
- Sorrendi (ordinális): nem csupán a skálaértékek azonos vagy nem azonos volta, hanem azok *sorrendisége* is vonatkoztatható a vizsgált egységekre. (pl. vizsgajegyek, iskolai végzettség, katonai rangok)
- Különbségi (intervallum): A skálaértékek *különbségei* is értelmezhetőek: mennyivel több/nagyobb/stb? A skála szerves tartozéka valamilyen mértékegység. (pl. hőmérséklet, naptári idő)
- Arány: Van természetes kezdőpontja. Két skálaérték egymáshoz viszonyított aránya meghatározható és értelmezhető. (pl. életkor, népesség)



Alapfogalmak: ismérvek és mérési skálák





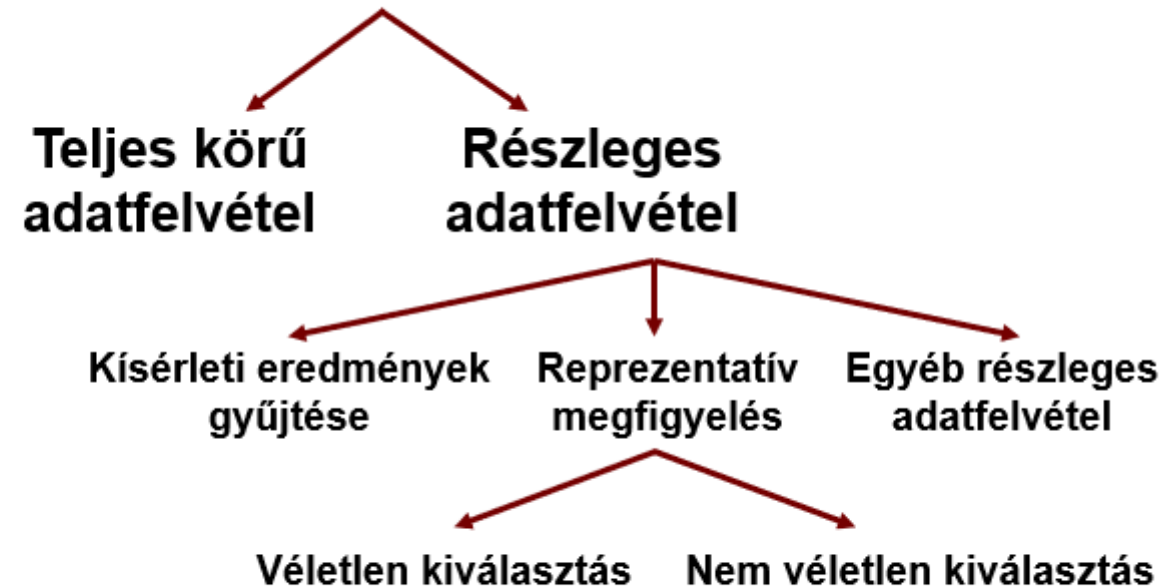
Adatfelvétel

1. alkalom





1. ADATFELVÉTELEK



2. ADMINISZTRATÍV NYILVÁNTARTÁSOK



Statisztikai alapműveletek

1. alkalom



Statisztikai alpműveletek



- 1) A sokaság nagyságának meghatározása
 - diszkrét - számlálás
 - folytonos - mérés
- 2) Összehasonlítás (két vagy több sokaság egészét jellemző adatok felsorolása, különbsége, hányadosa)
- 3) Osztályozás (ismérv szerinti tagolás, csoportosítás)
 - egy ismérv szerint: csoportosító sor
 - több ismérv szerint: kombinációs tábla

Csoportosító sor



A csoportosító sor általános sémája:

ahol:

C_i ismérvváltozat

f_i gyakoriság

N sokaság nagysága

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

Osztály	Egységek száma
C_1	f_1
C_2	f_2
\vdots	\vdots
C_i	f_i
\vdots	\vdots
C_k	f_k
Összesen	N

Csoportosító sor: példa



A magyar népesség nem szerint 2019. január 1. (fő)

Nem	f_i
Nő	5 096 935
Férfi	4 675 821
Összesen	9 772 756

Csoportosító sor: példa



A magyar népesség a település jellege szerint 2019. január 1. (fő)

Településjelleg	f_i
Budapest	1 752 286
Többi város	5 136 253
Község	2 884 217
Összesen	9 772 756

Csoportosító sor: példa



A magyar népesség korcsoport szerint 2019. január 1. (fő)

Életkor	f_i
0 - 14	1 421 739
15 - 64	6 461 058
65 -	1 889 959
Összesen	9 772 756

Kombinációs tábla



A sokaságot egyszerre több ismerv szerint osztályozzuk
(statisztikai sorok összekapcsolása)

ahol:

$C_i^X ; C_j^Y$ ismérvváltozatok

$f_{i.} ; f_{.j}$ peremgyakoriságok

$r ; c$ ismérvváltozatok száma

f_{ij} az X ismerv i-edik és az Y ismerv j-edik változatához tartozó gyakoriság

X ismerv szerinti osztályok	Y ismerv szerinti osztályok						
	C_1^Y	C_2^Y	\dots	C_j^Y	\dots	C_c^Y	\sum_j
C_1^X	f_{11}	f_{12}	\dots	f_{1j}	\dots	f_{1c}	$f_{1.}$
C_2^X	f_{21}	f_{22}	\dots	f_{2j}	\dots	f_{2c}	$f_{2.}$
\vdots	\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
C_i^X	f_{i1}	f_{i2}	\dots	f_{ij}	\dots	f_{ic}	$f_{i.}$
\vdots	\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
C_r^X	f_{r1}	f_{r2}	\dots	f_{rj}	\dots	f_{rc}	$f_{r.}$
\sum_i	$f_{.1}$	$f_{.2}$	\dots	$f_{.j}$	\dots	$f_{.c}$	N

Kombinációs tábla: példa



A magyar népesség korcsoport és nem szerint
2019. január 1. (fő)

Életkor	Nem		
	Férfi	Nő	Összesen
0 - 14	729 954	691 785	1 421 739
15 - 64	3 228 776	3 232 282	6 461 058
65 -	717 091	1 172 868	1 889 959
Összesen	4 675 821	5 096 935	9 772 756

Kombinációs tábla: példa



A magyar népesség korcsoport szerint a vizsgált években, fő, (január 1.)

Életkor	Év	
	1980	2019
0 - 14	2 341 173	1 421 739
15 - 64	6 918 844	6 461 058
65 -	1 449 446	1 889 959
Összesen	10 709 463	9 772 756

Feladat



- Nyissa meg az „alkalom_01.xlsx” Excel fájlt, nézze meg az adatbázist és hajtsa végre a megadott feladatokat!

Házi feladat



Menjen a Core online-könyvek „Doing Economics”, 1. fejezetéhez: Measuring Climate Change.

Olvassa el a bevezetőt:

<https://www.core-econ.org/doing-economics/book/text/01-01.html>

Készítse el az 1.1 fejezet feladatait és adja le a megfelelő Excel fileokat a Moodle felületén.

Azokon a helyeken, ahol a szöveg ezt kéri (pl. hasonlítsa össze vagy javasoljon, stb.) ott írjon 1-2 mondatot az Excel egyik, jól látható cellájába!

<https://www.core-econ.org/doing-economics/book/text/01-02.html>

Köszönöm a figyelmet!





Corvinus



2. alkalom: Adattípusok, leíró statisztikák

Budapesti Corvinus Egyetem



Budapesti Corvinus Egyetem
Corvinus University of Budapest

Tartalom



Adattípusok: keresztmetszet, idősor és panel

Adatok grafikus ábrázolása

Minőségi és mennyiségi adatok

Adatok közötti összefüggés

Leíró statisztikák



Adattípusok: (1) idősoros adatok



Idősoros adatok

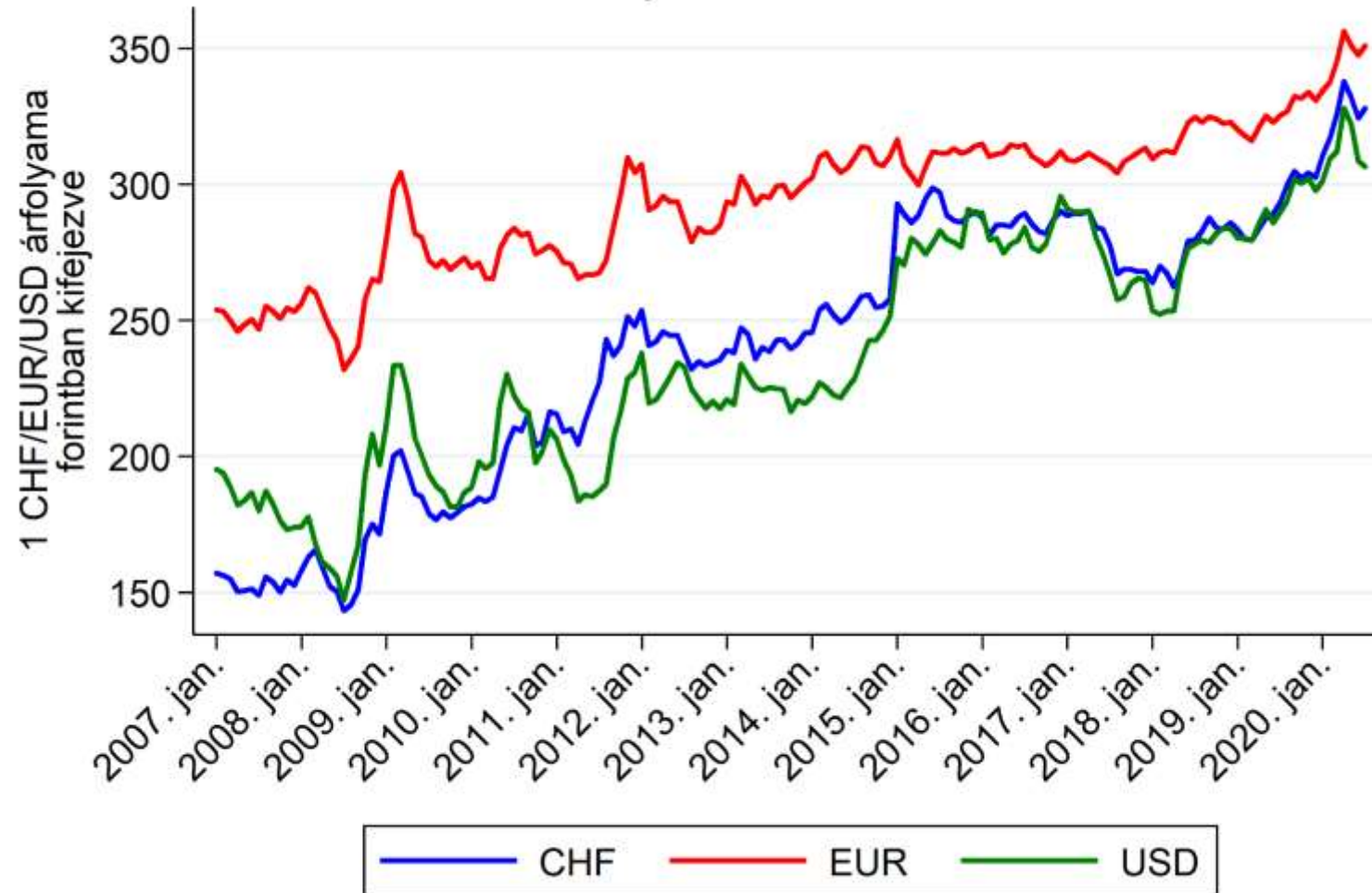


- Időben rendezett változók, sorrend számít.
- Megfigyelési gyakoriság: évi, havi, heti, napi, sőt még gyakoribb – példák?
- Jelölés: Y_t
- Pl.: a Ft/CHF árfolyam, a GDP, a népességszám alakulása.
- A grafikus ábrázolás segít tömören áttekinteni az adatokat.

Adatok időbeni alakulásának grafikus ábrázolása



Valuták árfolyama a forinttal szemben

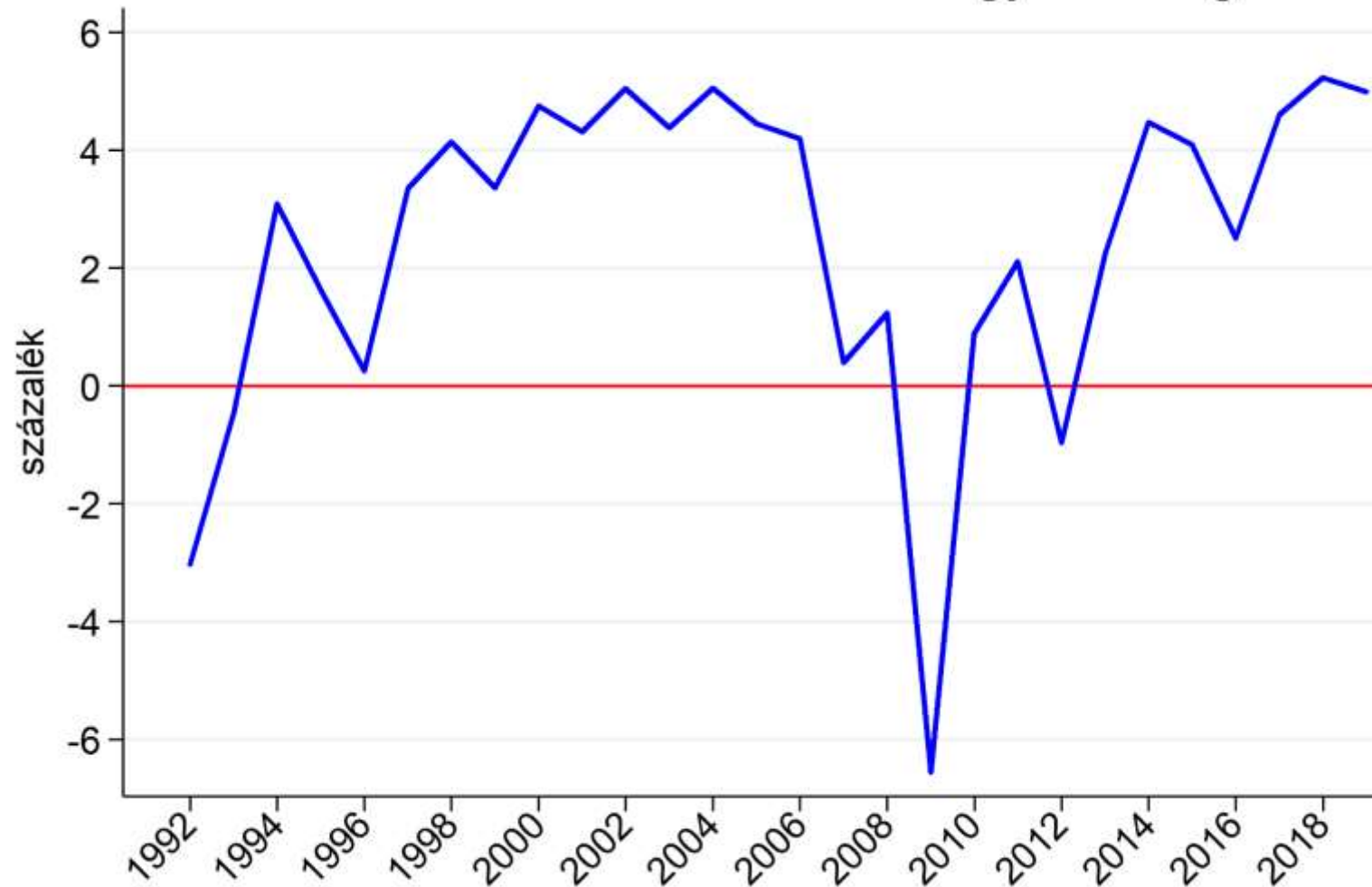


Forrás: MNB. Megjegyzés: Havi átlagos árfolyamok.

Adatok időbeni alakulásának grafikus ábrázolása



A GDP/fő növekedési üteme Magyarországon



Forrás: WB. Megjegyzés: Konstans áras GDP.



Adattípusok: (2)

keresztmetszeti adatok

2. alkalom



Keresztmetszeti adatok



- Egy adott időpontban a gazdaság szereplőiből (egyén, vállalat, ország stb.) vett minta – pillanatfelvétel az időben minket érdeklő szereplőkről.
- Sorrend nem számít.
- Jelölés: Y_i
- Sokszor véletlen minta: például a bérek érdekelnek egy adott szektorban, akkor véletlenszerűen kiválasztok X személyt az adott szektorból és megnézem a bérüket. Általában lehetetlen mindenkit megfigyelni.
- Máskor minden egység megfigyelhető: például országok esetén, hogy diktatúrák-e vagy az urbanizációs szintjük.

Keresztmetszeti adatok ábrázolása: hisztogram

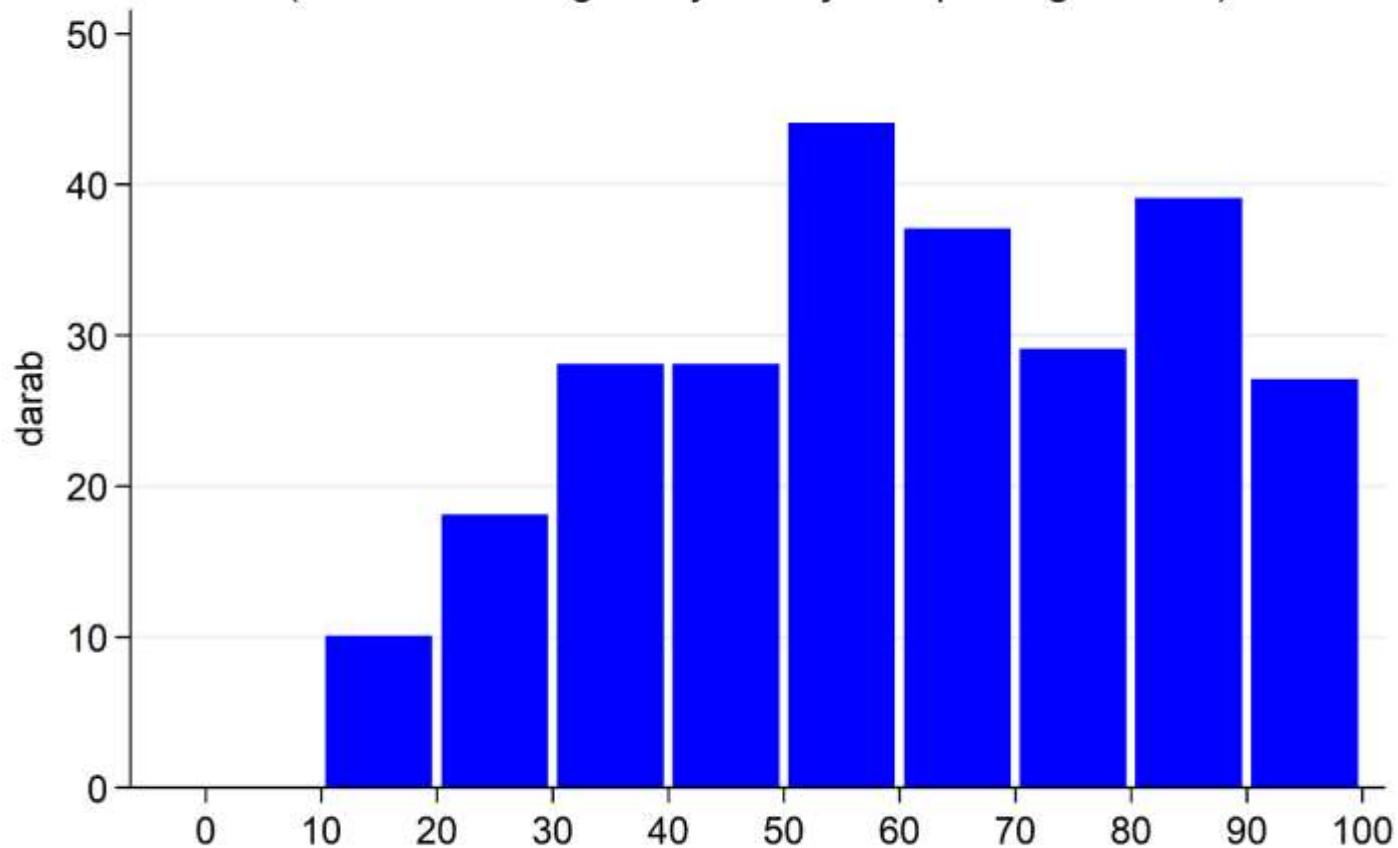


- Hisztogram: diszjunkt halmazokba tartozó elemeket összeszámoló függvény.
- Példa: egy főre jutó jövedelem megoszlása (eloszlása).
- Egyenlő osztályközök (rekesznagyság) – Excelben célszerű megadni adatoktól függően – hisztogram érzékeny lehet a rekesznagyságra.
- Gyakoriság egyes osztályközökben.

Keresztmetszeti adatok: eloszlás grafikus ábrázolása (hisztogram)



Urbanizáció foka a világban
(Városi lakosság aránya a teljes népességén belül)



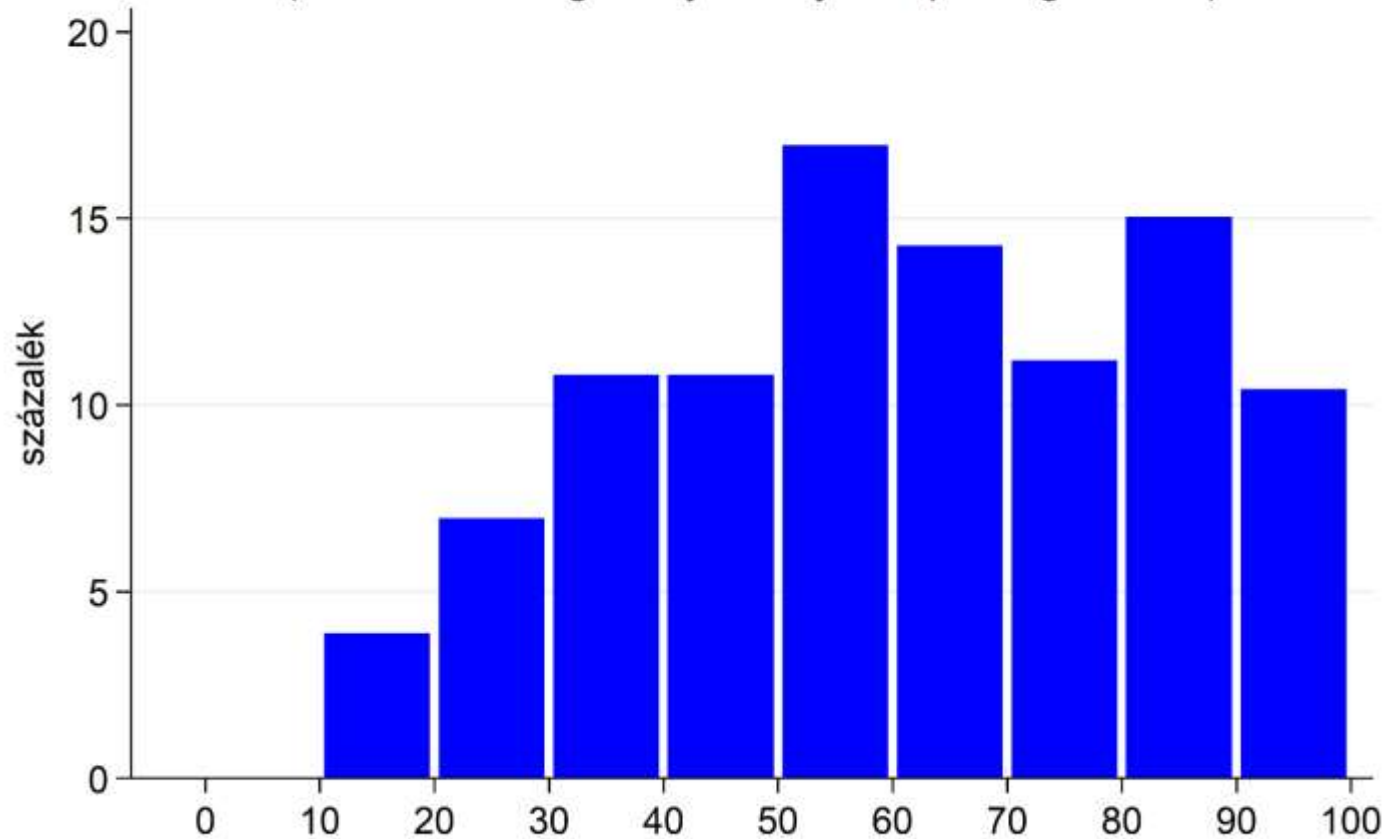
Osztályköz	Gyakoriság (darab)
0-10	0
10-20	10
20-30	18
30-40	28
40-50	28
50-60	44
60-70	37
70-80	29
80-90	39
90-	27
Összesen	260

Forrás: Világbank. Megjegyzés: 2019-re vonatkozó adatok.

Keresztmetszeti adatok: eloszlás grafikus ábrázolása (hisztogram)



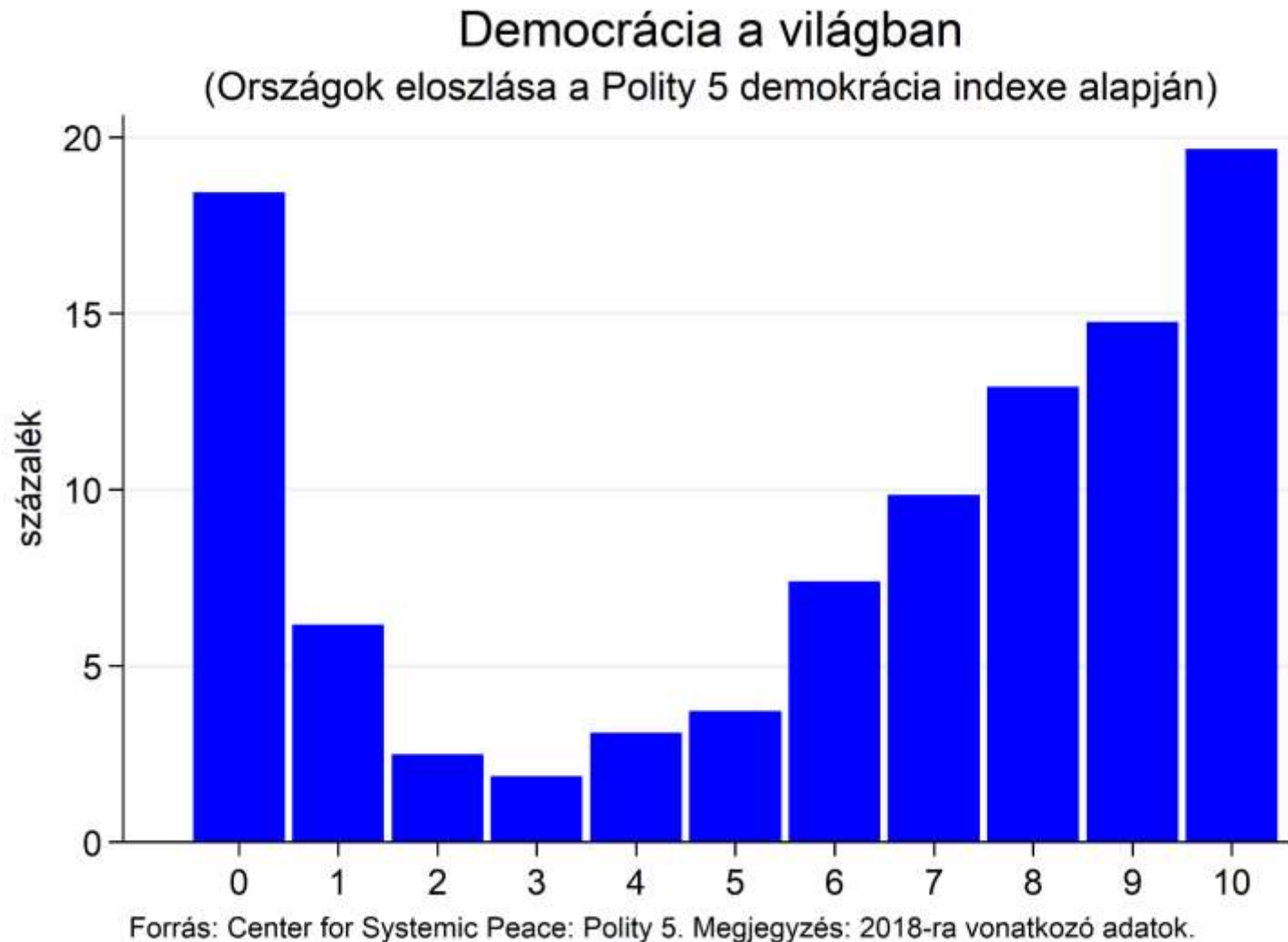
Urbanizáció foka a világban
(Városi lakosság aránya a teljes népességen belül)



Forrás: Világbank. Megjegyzés: 2019-re vonatkozó adatok.

Osztályköz	Gyakoriság (százalék)
0-10	0,0
10-20	3,8
20-30	6,9
30-40	10,8
40-50	10,8
50-60	16,9
60-70	14,2
70-80	11,2
80-90	15,0
90-	10,4
Összesen	100

Keresztmetszeti adatok: eloszlás grafikus ábrázolása (hisztogram)



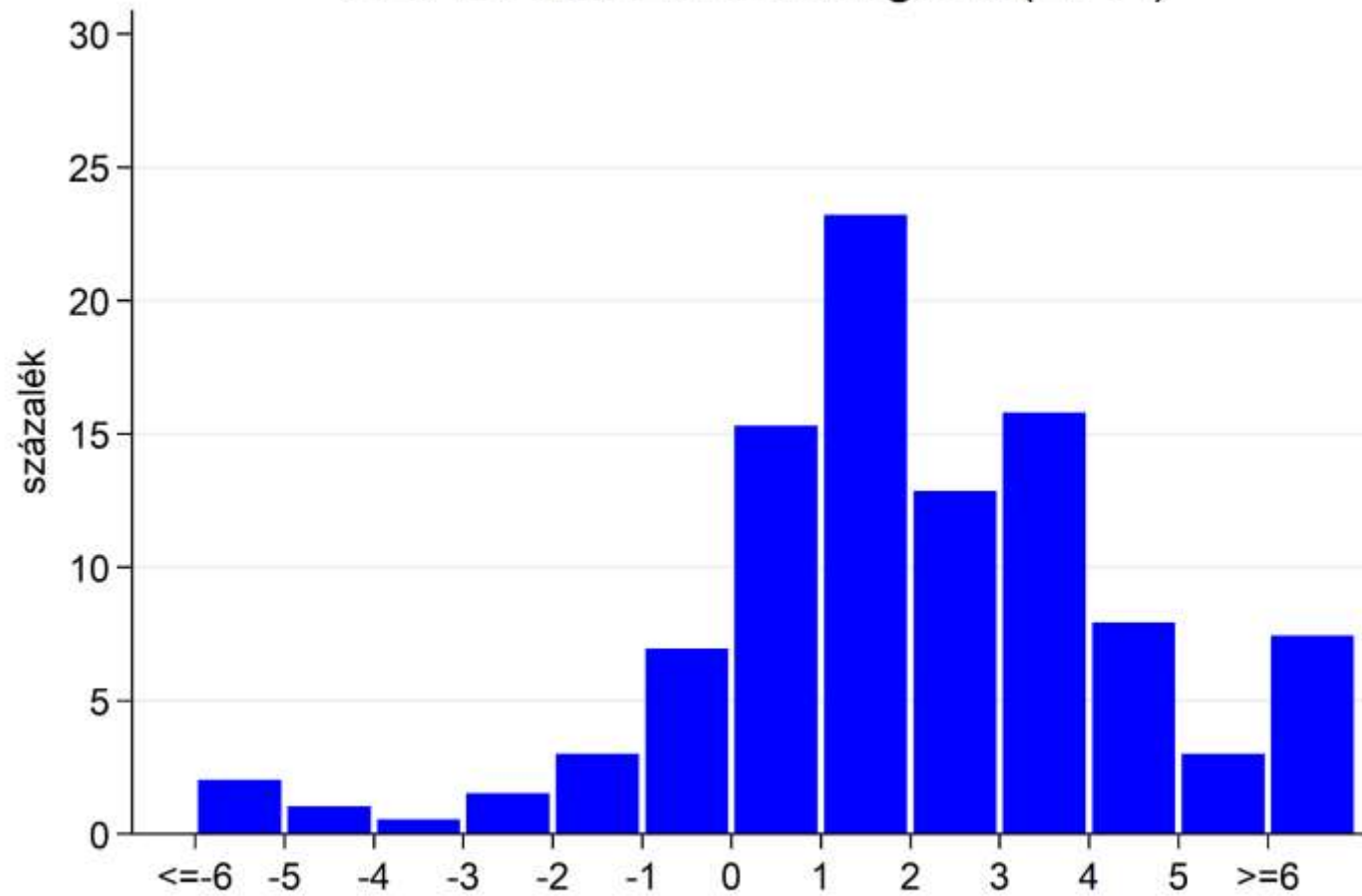
Két pólusú eloszlás.

Egy másik lehetséges ábrázolás:
<https://ourworldindata.org/grapher/political-regime-updated2016>

Keresztmetszeti adatok: eloszlás grafikus ábrázolása (hisztogram)



GDP/fő változása a világban (2014)



Forrás: Világbank.



Adattípusok: (3) panel adatok



Panel adatok



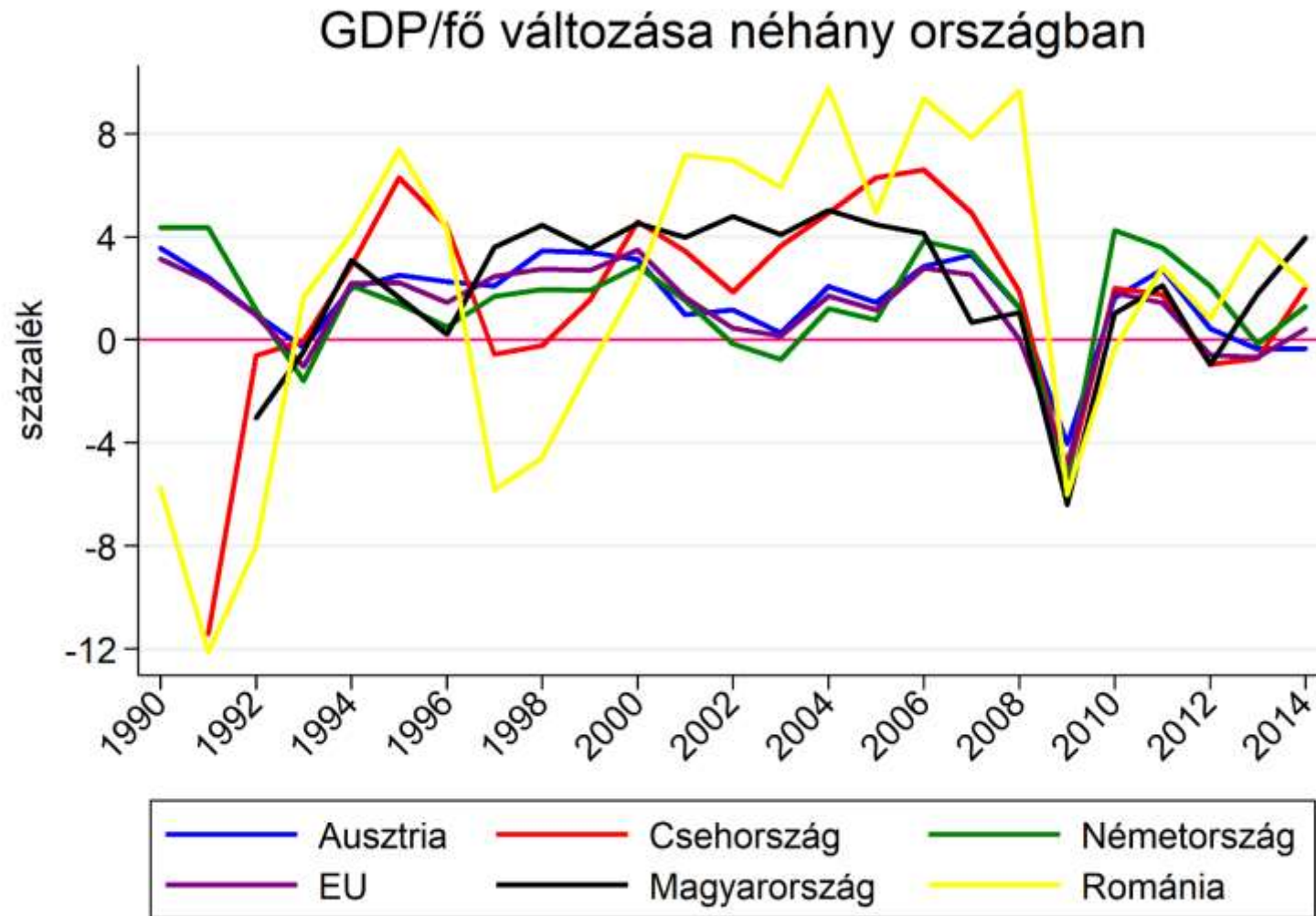
- Idősoros + keresztmetszeti együttesen.
- Keresztmetszeti mintáról megfigyelés több időpontban.
- Jelölés: Y_{it}
- Összesen $T \times N$ megfigyelés, ahol N a megfigyelt egységek (pl. egyének, országok) száma.
- Példák: GDP európai országokban, egyéni keresetek alakulása, urbanizáció alakulása a nagyvilágban.

Panel adatok ábrázolása



- Ha kevés megfigyelt egység, akkor lehet ábrázolni normálisan. A következő ábrán majd látszik, hogy ilyenkor sem könnyű átlátni az ábrát.
- Ha azonban sok megfigyelt egység van, akkor nem lehet szépen ábrázolni.

Panel adatok grafikus ábrázolása



Forrás: Világbank

Feladat



- Nyissa meg a Világbank online adatbázisát: <https://data.worldbank.org/>
- Keressen országokra vonatkozó *teljes népesség* adatot: [Population, total](#)
- Az adatokat töltsse le egy Excel fájlba!
- Hozzon létre a letöltött Excel fájlban három új fület *idősor*, *keresztmetszet*, és *panel* néven, majd a három föltre hozzon létre egy-egy adatbázist, amely az adott adatbázistípusba tartozik!



Szint és változás

2. alkalom



Adatok időbeni alakulása: szint vs. változás



- Sokszor nem a szint, hanem a változás érdekel minket. Szintet változással nem érdemes összehasonlítani.
- Szintből egyszerű növekedési rátát (változást) számolni.

$$\% \text{ változás} = \frac{(Y_{t+1} - Y_t)}{Y_t} \cdot 100$$

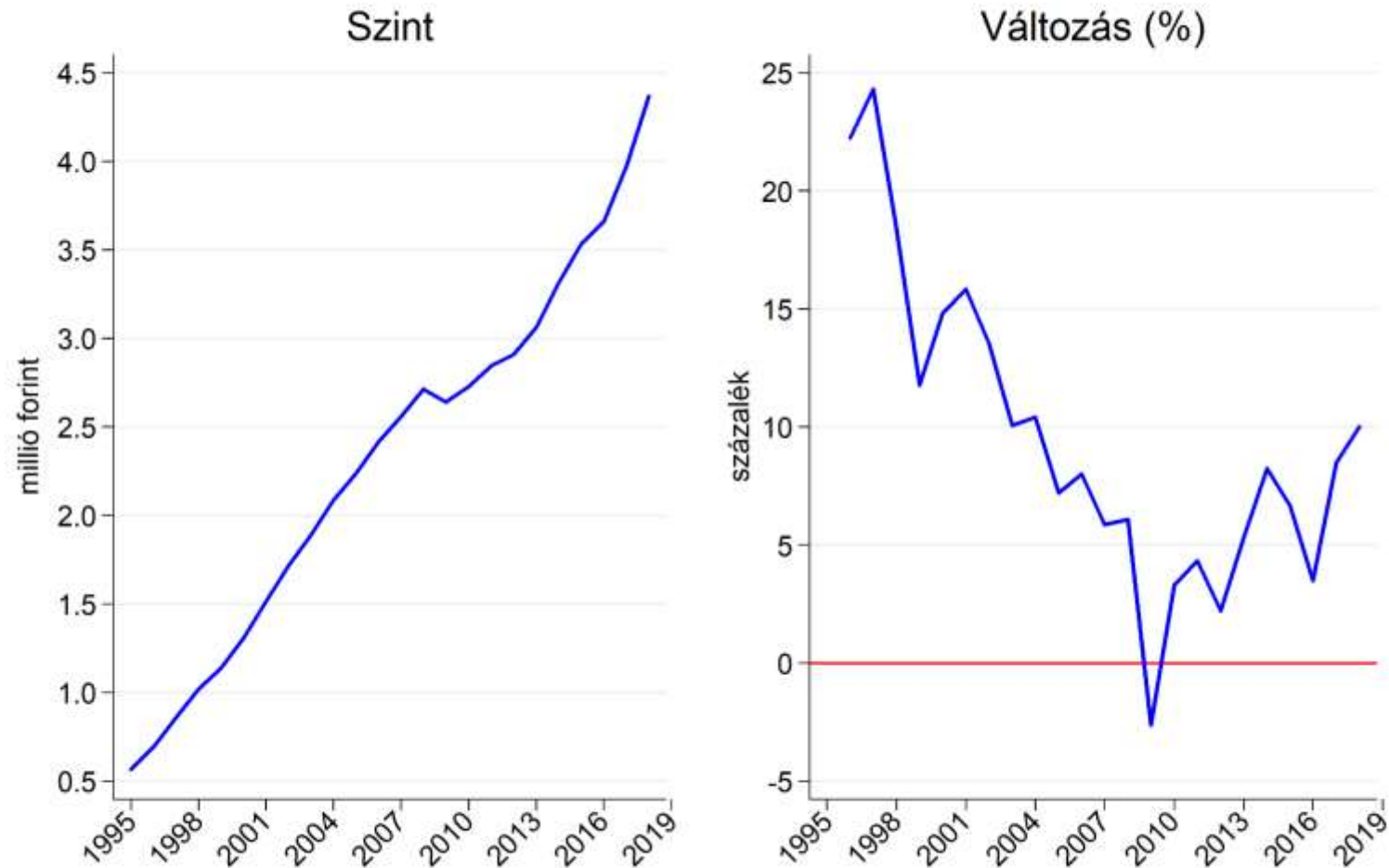
Feladat



- Nyissa meg a Központi Statisztikai Hivatal online adatbázisát:
<http://www.ksh.hu/>
- Keressen Magyarországra vonatkozó *GDP/fő* adatot az „Adatok, kiadványok” menüpont „Összefoglaló táblák (STADAT)” pontjában!
- Az adatokat töltsse le egy Excel fájlba! A HUF-ban mért adatokat számolja át millió HUF-ra és ezt használja a továbbiakban!
- Ábrázolja a magyar *GDP/fő* értéket időben, a lehető leghosszabb intervallumon!
- Számolja ki a *GDP/fő* éves növekedési rátáját! Ábrázolja is azt!
- Hasonlítsa össze a két ábrát!



A GDP/fő időbeli alakulása Magyarországon





Adatok közötti összefüggés grafikus bemutatása

2. alkalom



Adatok közötti összefüggés bemutatása



- Két változó összefüggését úgy vizsgálhatjuk, hogy megnézzük, hogy együttmozognak-e.
- Példák:
 - A jegybanki alapkamat tényleg hat az inflációra?
 - Az országok közötti beruházásban tapasztalt különbségek magyarázzák a GDP-változásbeli különbségeket?
 - Magasabb-e a GDP ott, ahol magasabb a felsőfokú végzettségűek aránya?
 - Diktatúrákban magasabb-e a városiasodás?

Adatok közötti összefüggés bemutatása

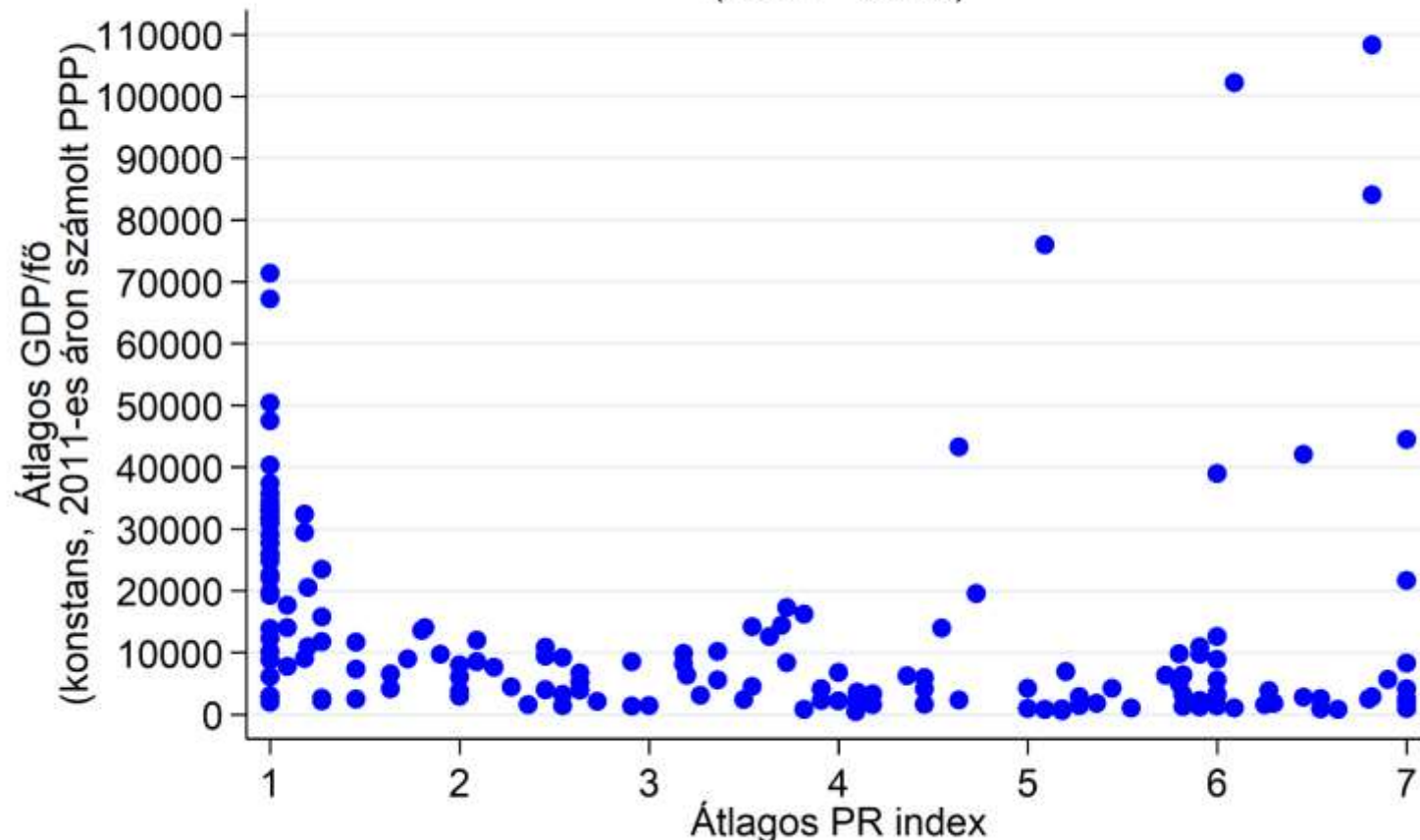


- Figyelem: együttmozgás nem feltétlenül jelent oksági viszonyt!
 - Pl.: csokoládéfogyasztás és Nobel-díjak száma: <https://www.sciencedirect.com/science/article/pii/S2590291120300711>
 - Pl.: Gólyák és születésszám: <https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9639.00013>
 - Pl.: Elhízott politikusok és korrupció: <https://onlinelibrary.wiley.com/doi/abs/10.1111/ecot.12259>
 - További példák: <https://www.tylervigen.com/spurious-correlations>
- Az összefüggés jellemzően nem determinisztikus. A tendenciák érdekelnek minket. Sokszor vannak kiugró értékek (outlierek), amik nem illenek a tendenciába, ami nem baj. Azokból is sokat lehet tanulni.

Adatok közötti összefüggés bemutatása: pontdiagram



Átlagos GDP/fő és politikai jogok (PR) index értékek
(1990 - 2000)



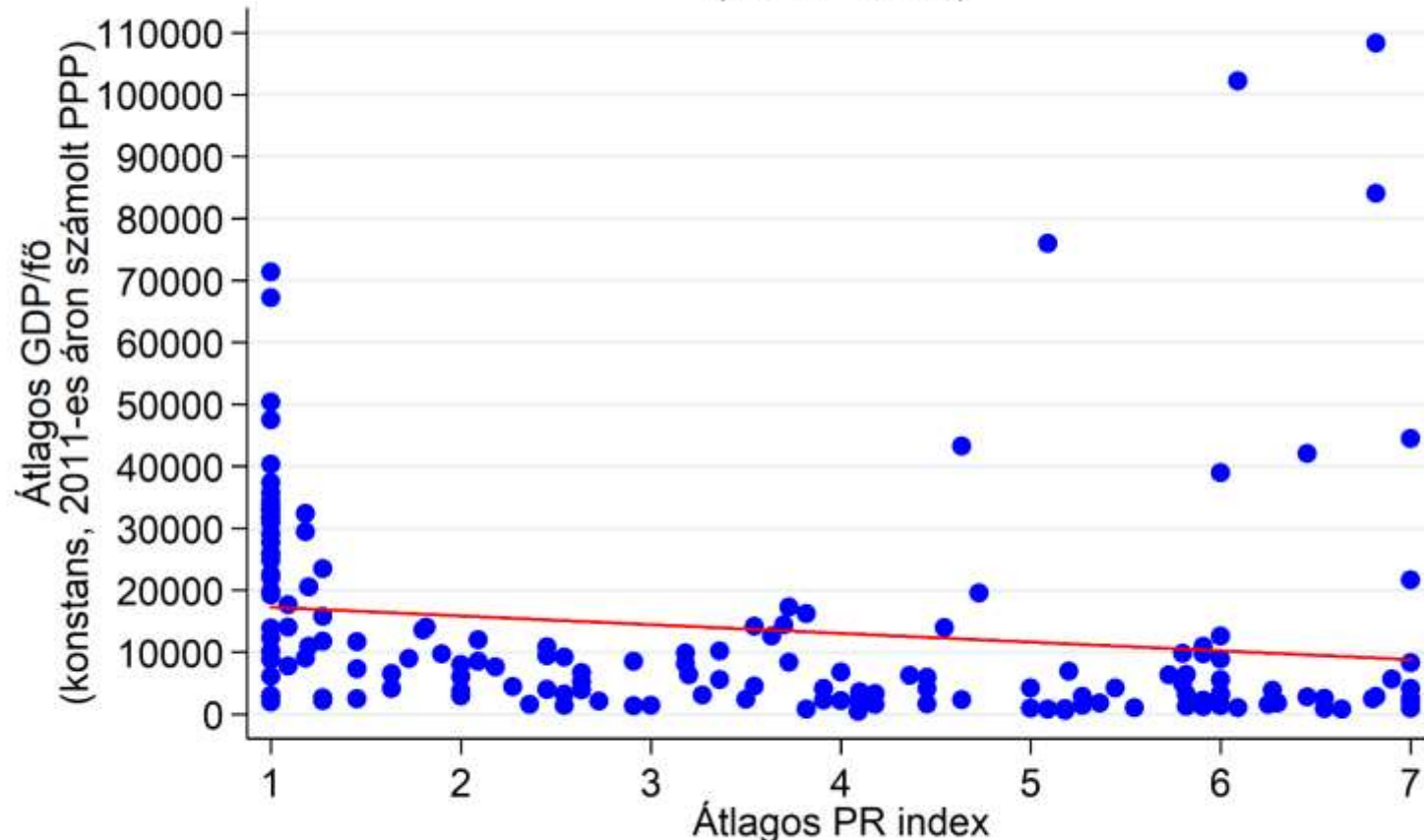
- Legalapvetőbb módja az együttmozgás megvizsgálásának a pontdiagram.

Forrás: Freedom House és Világbank.
Megjegyzés: A magasabb PR érték gyengébb demokráciát jelent.

Adatok közötti összefüggés bemutatása: pontdiagram



Átlagos GDP/fő és politikai jogok (PR) index értékek
(1990 - 2000)



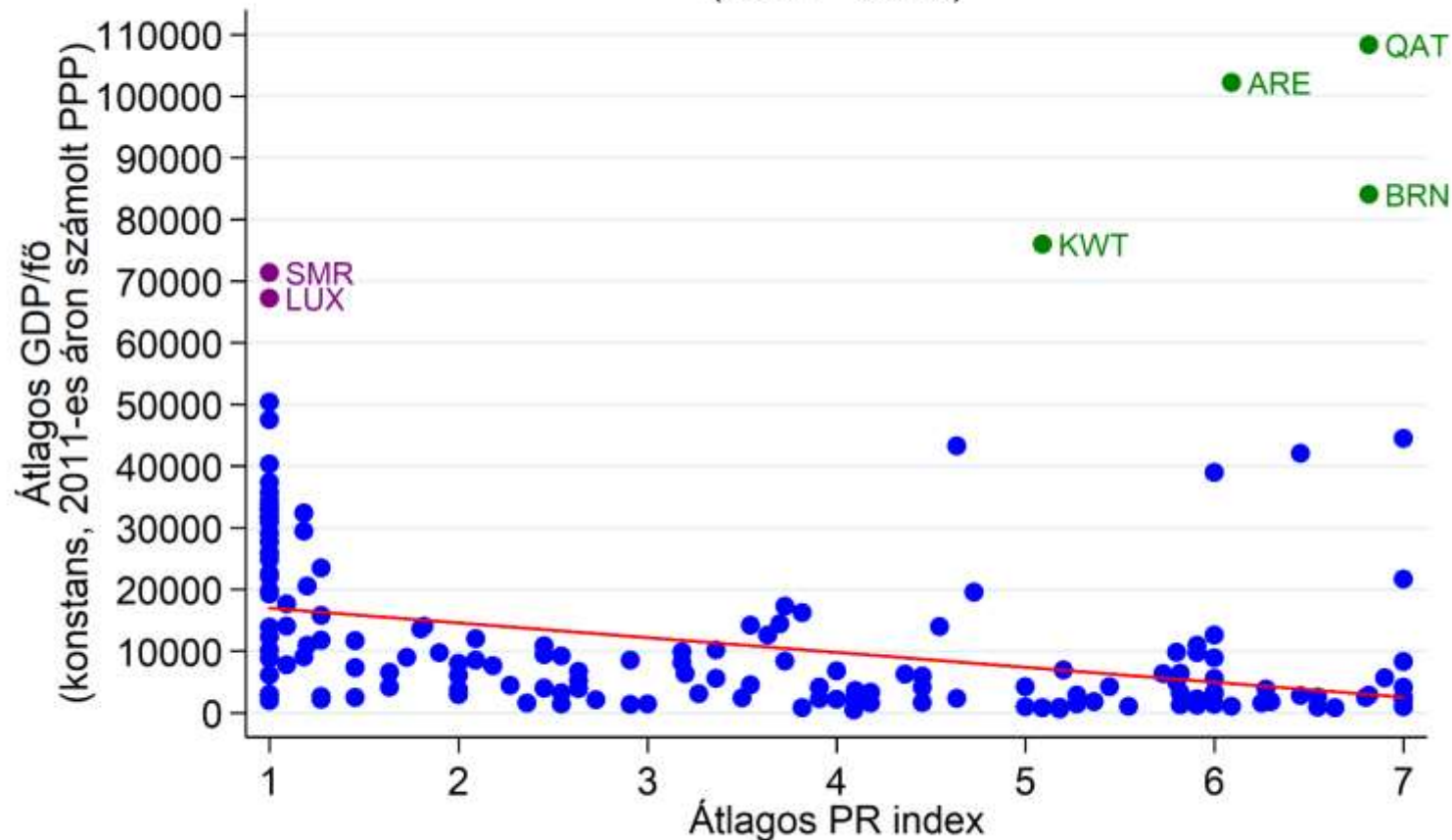
Forrás: Freedom House és Világbank.
Megjegyzés: A magasabb PR érték gyengébb demokráciát jelent.

- Egyenes illesztése segíthet a kapcsolat „láthatóvá válásában”.
- Igazán jó megoldás: számszerű leírás (kapcsolatvizsgálat) – később.

Adatok közötti összefüggés bemutatása: pontdiagram



Átlagos GDP/fő és politikai jogok (PR) index értékek
(1990 - 2000)



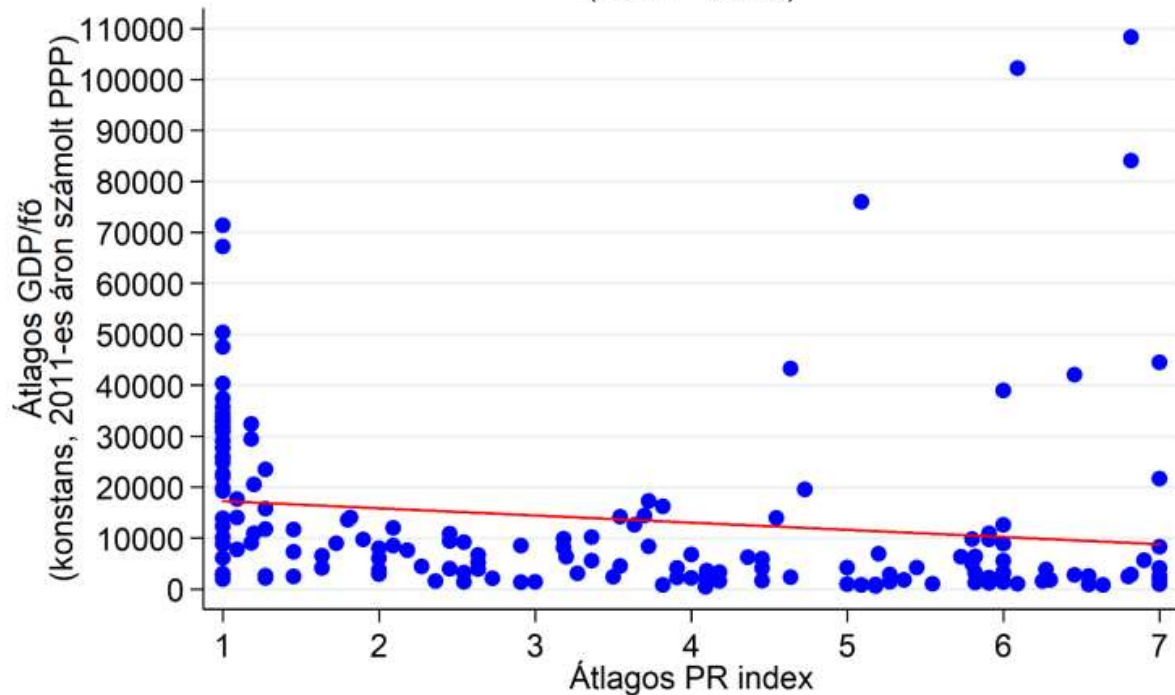
- Outlier-ek!

Forrás: Freedom House és Világbank.
Megjegyzés: A magasabb PR érték gyengébb demokráciát jelent.

Adatok közötti összefüggés bemutatása: pontdiagram

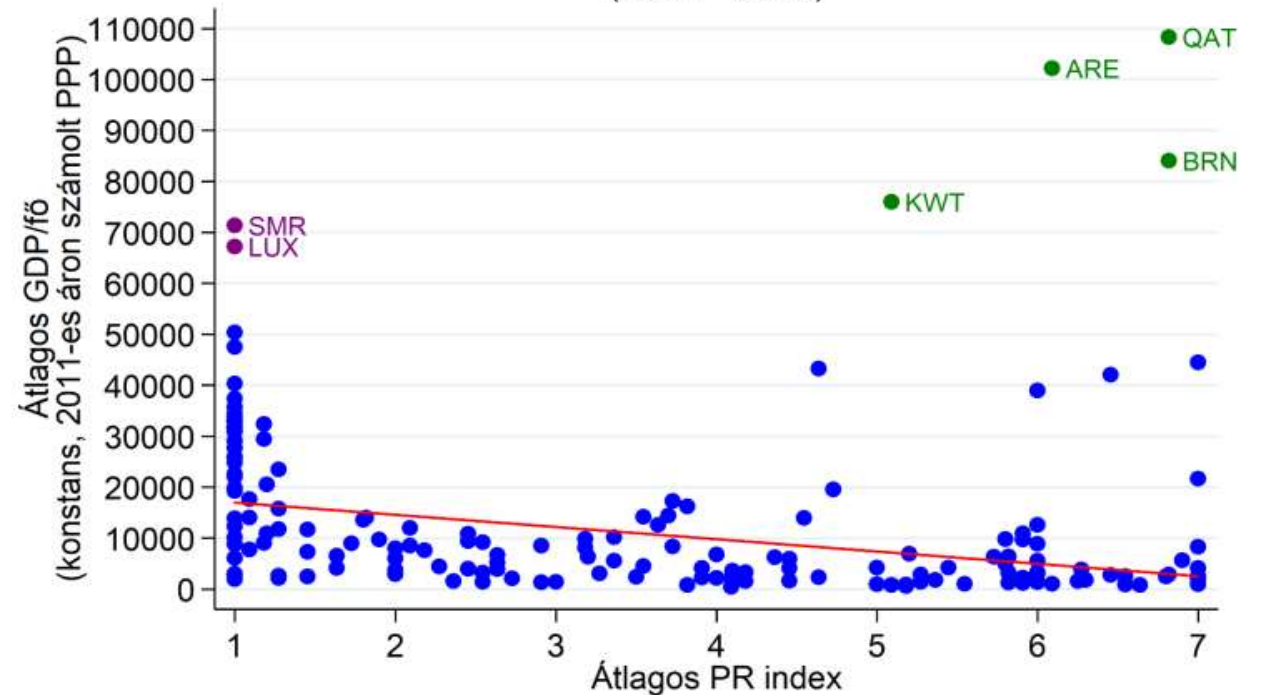


Átlagos GDP/fő és politikai jogok (PR) index értékek
(1990 - 2000)



Forrás: Freedom House és Világbank.
Megjegyzés: A magasabb PR érték gyengébb demokráciát jelent.

Átlagos GDP/fő és politikai jogok (PR) index értékek
(1990 - 2000)



Forrás: Freedom House és Világbank.
Megjegyzés: A magasabb PR érték gyengébb demokráciát jelent.



Leíró statisztikák

2. alkalom



Leíró statisztikák



Leíró statisztika: számszerűen és tömören összefoglalni változó jellemzőit

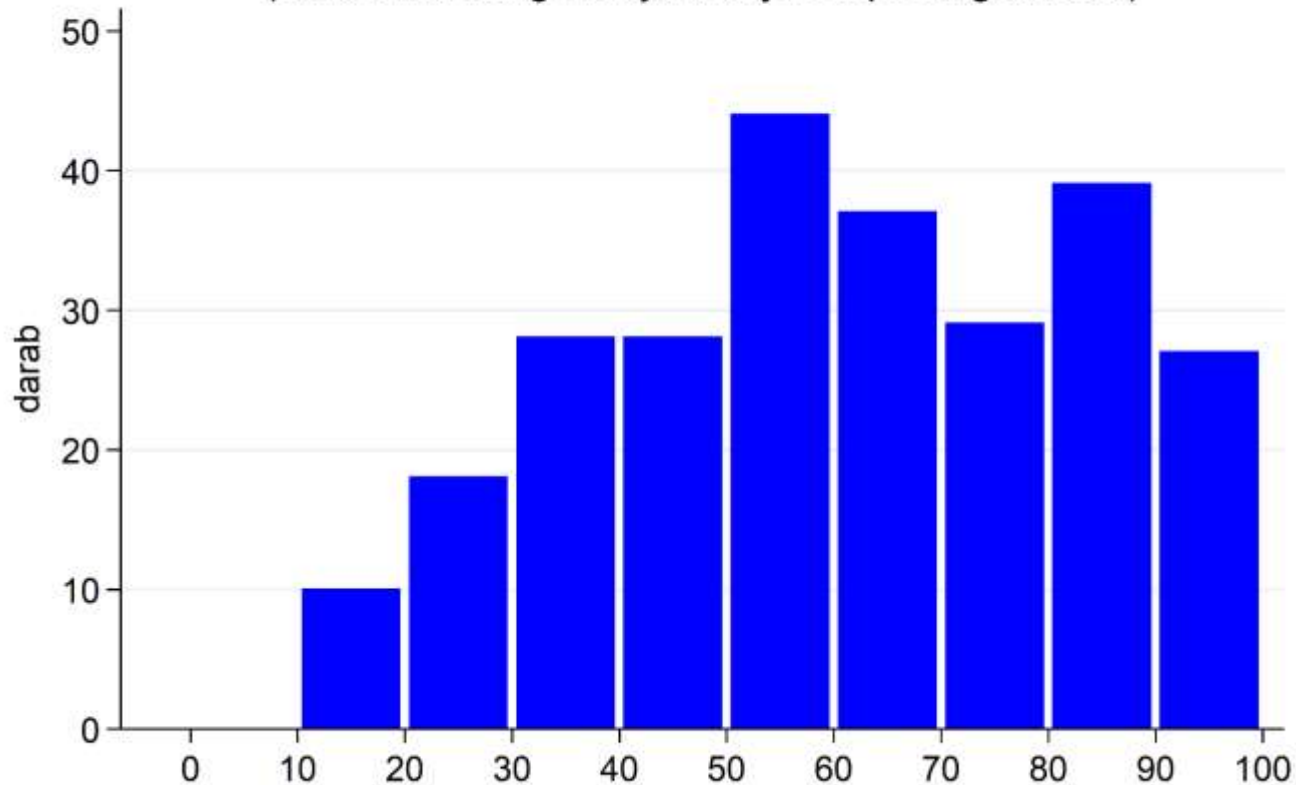
- Szintje? – átlag, medián (percentilisek), módusz
- Változékonysága? – szórás, terjedelem, interkvartilis terjedelem

Leíró statisztikák: átlag



- N: minta elemszáma, megfigyelések száma

Urbanizáció foka a világban
(Városi lakosság aránya a teljes népességen belül)



Forrás: Világbank. Megjegyzés: 2019-re vonatkozó adatok.

$$\bar{Y} = \frac{\sum_{i=1}^N Y_i}{N}$$

Átlag = 60,4

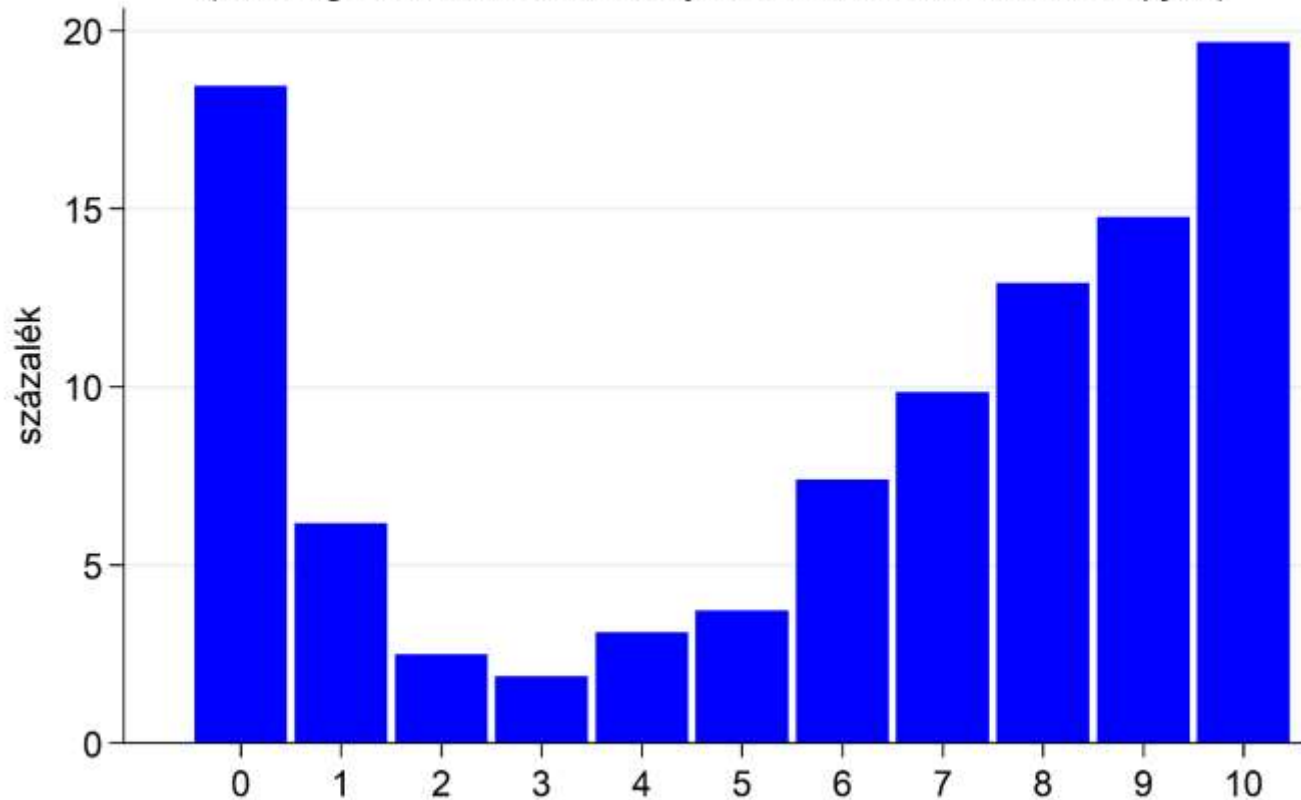
Leíró statisztikák: átlag



- Probléma: Két csúcsú eloszlásnál nem biztos, hogy informatív.

Democrácia a világban

(Országok eloszlása a Polity 5 demokrácia indexe alapján)

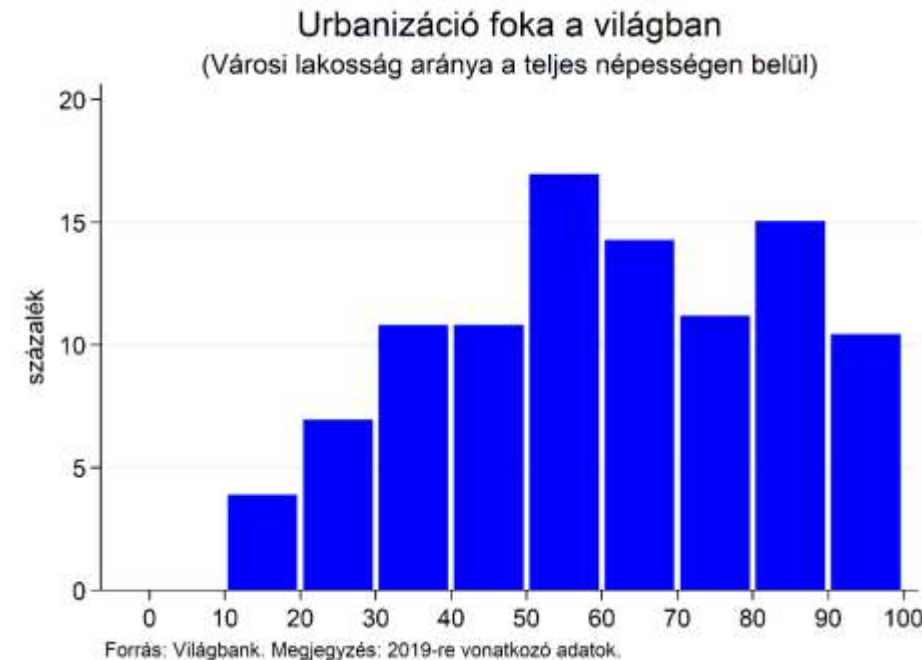
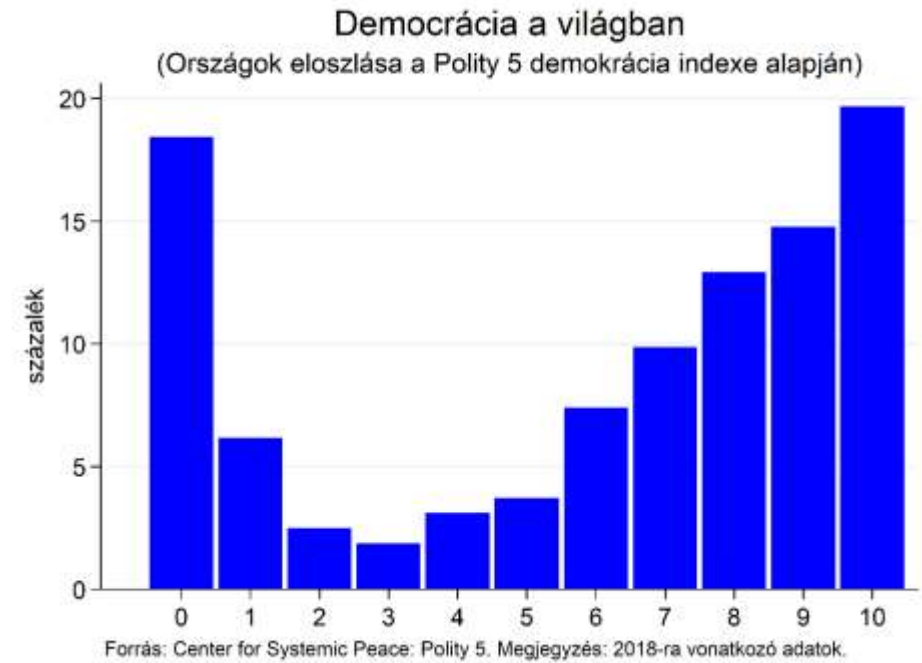


Átlag = 5,9

Forrás: Center for Systemic Peace: Polity 5. Megjegyzés: 2018-ra vonatkozó adatok.

Leíró statisztikák: módusz

- Módusz = leggyakoribb érték
- Probléma: nem mindig létezik (pl. minden értékből egy), ill. több módusz is lehet
- Megoldás lehet: hisztogram legmagasabb pontja (függ osztályközöktől) – osztályköz közepe
- Példa:
 - Demokrácia: 10 (vö.: átlag=5,9)
 - Urbanizáció: 50-60 rekesz vagy 55% (vö.: átlag=60,4)



Leíró statisztikák: medián, percentilisek



- Medián: középső érték – megfigyelések fele alatta, fele felette van – a demokrácia foka esetén a medián 7
- X-edik percentilis: megfigyelések X%-a kisebb értéket vesz fel
- Kvartilis / Decilis: negyedeli / tizedeli az adatokat
 - 1. kvartilis: 25% alatta, 2. kvartilis = medián
 - 1. decilis: 10 % alatta, 5. decilis = medián

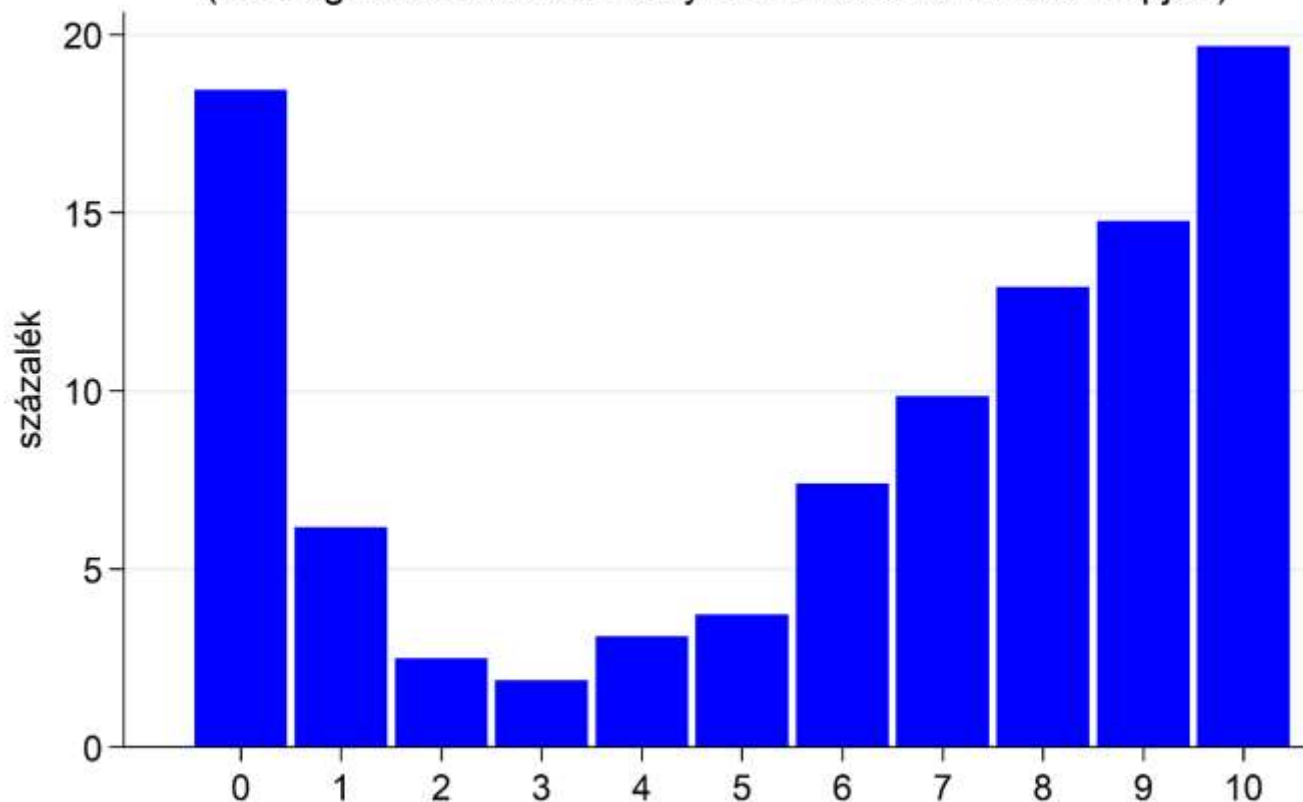
Leíró statisztikák: medián



Medián = 7

Democrácia a világban

(Országok eloszlása a Polity 5 demokrácia indexe alapján)



Forrás: Center for Systemic Peace: Polity 5. Megjegyzés: 2018-ra vonatkozó adatok.

Érték	Gyakoriság (%)	Kumulált gyakoriság (%)
0	18,4	18,4
1	6,1	24,5
2	2,5	27,0
3	1,8	28,8
4	3,1	31,9
5	3,7	35,6
6	7,4	43,0
7	9,8	52,8
8	12,9	65,7
9	14,7	80,4
10	19,6	100,0

Átlag, módusz, medián - összefoglalva



- **Átlag:**
 - Bármely adathalmazból meghatározható és minden adatot felhasznál.
 - Érzékeny a szélsőségesen nagy vagy kicsi értékekre.
- **Módusz:**
 - Nem mindig határozható meg egyértelműen, sőt, nem is mindig létezik.
 - Nem függ közvetlenül sem az összes értéktől, sem a kiugró értékektől.
 - Meghatározható nominális skálán mért adatok esetében is.
- **Medián:**
 - Mindig egyértelműen meghatározható.
 - Ha a rangsorban nagyon sok egyforma ismértérték szerepel, akkor nem tanácsos használni.
 - Nem függ közvetlenül sem az összes értéktől, sem a kiugró értékektől.

Leíró statisztikák: átlag, módusz, medián

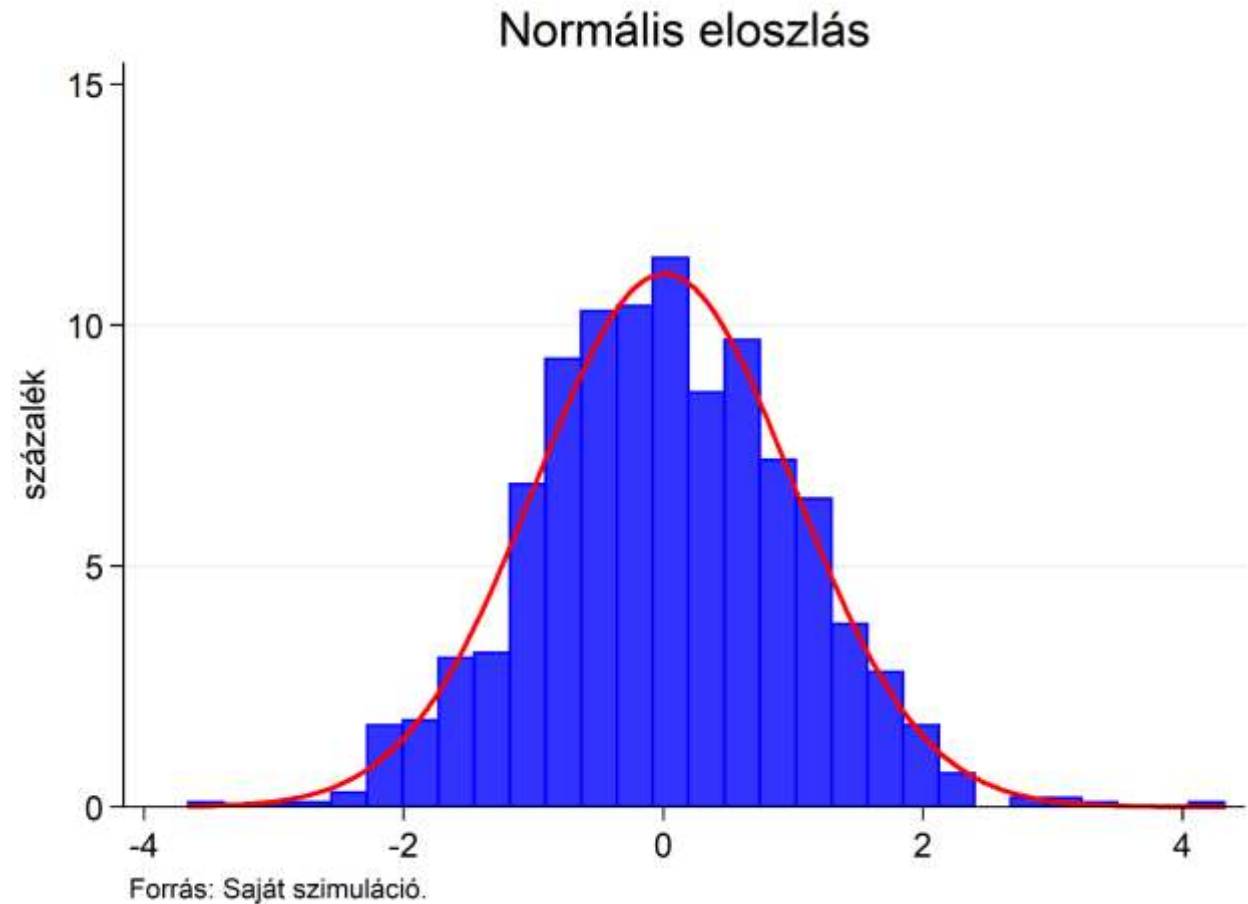


- A változó jellemző értékéről adnak számot különbözőképpen. (Central tendencies.)
- Egymáshoz viszonyított helyzetük (főleg az átlag és a medián viszonya) a változó eloszlásáról nyújt képet.
- Adatelemzéskor érdemes mindegyikre figyelni. Ne ragadjunk le az átlagnál, hiszen láttuk, hogy az átlag nem mindig jellemző érték.
- Normális eloszlásnál (majd lesz róla szó később) egybeesnek.

Átlag, módusz és medián normális eloszlás esetén



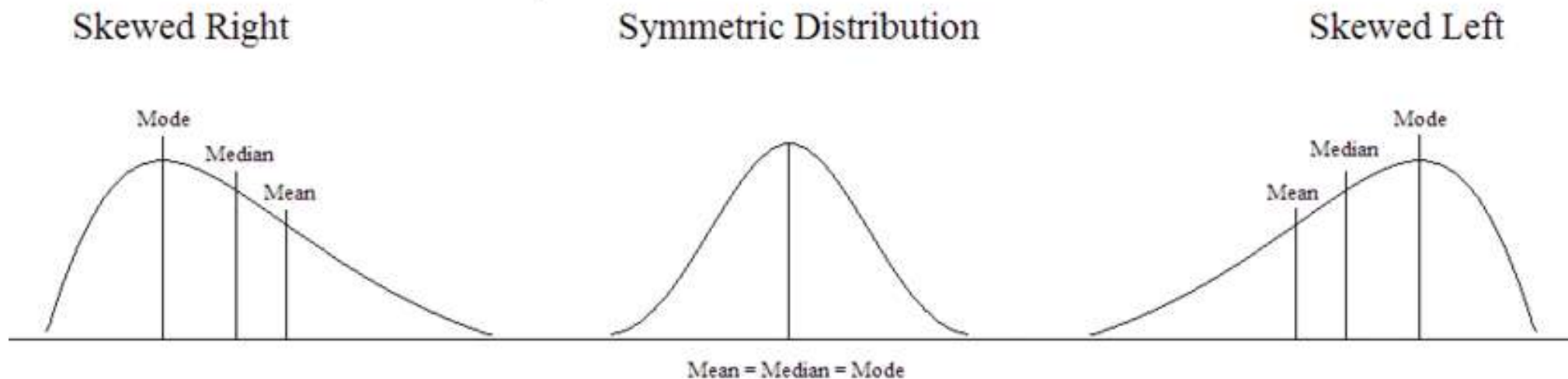
- Normális eloszlás esetén egybeesnek.
- Példa: emberek magassága, vérnyomás, zh-n szerzett eredmények.



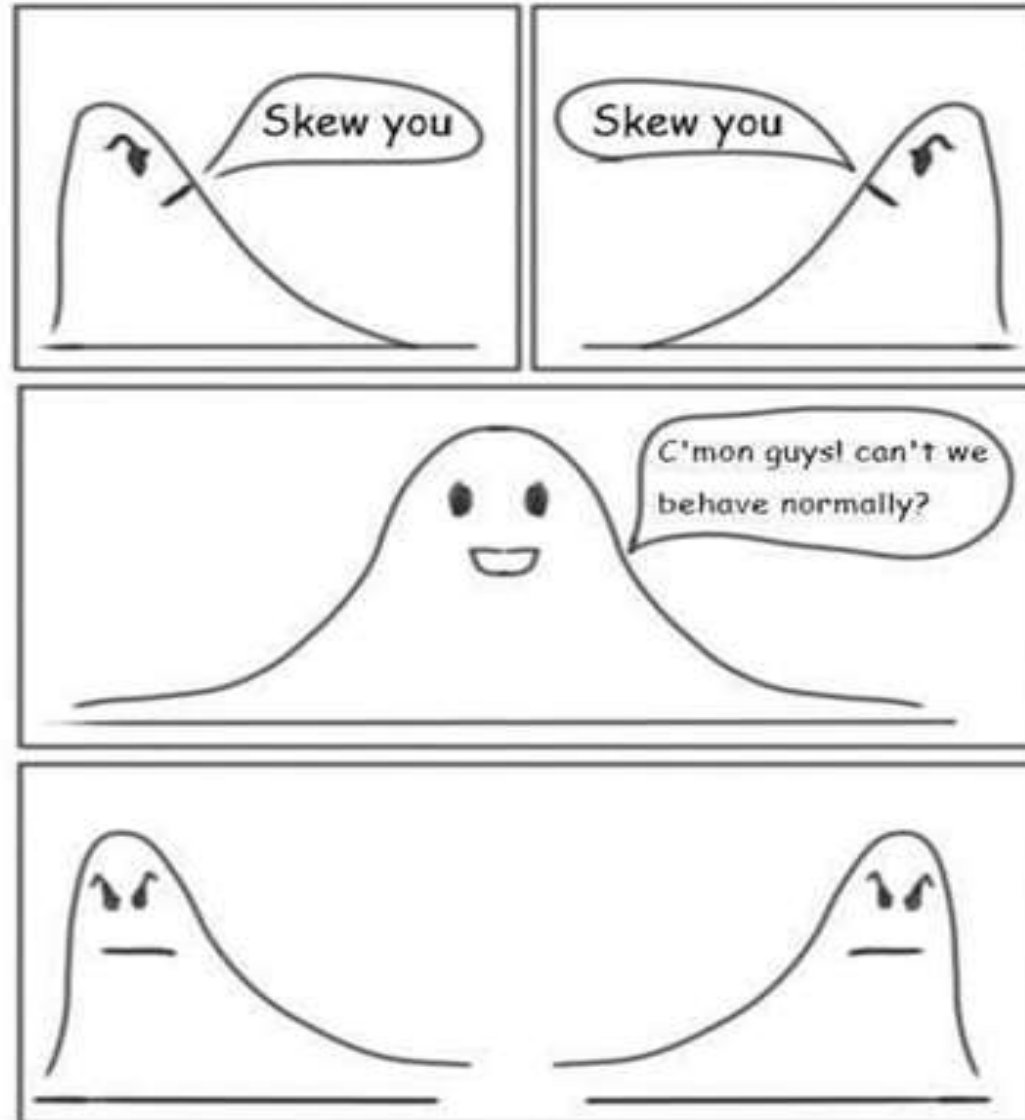
Leíró statisztikák: ferdeség (skewness)



- Szimmetrikus: átlag = medián = módusz
- Pozitív ferdeség (jobbra hosszan elnyúló): módusz < medián < átlag
- Negatív ferdeség (balra hosszan elnyúló): módusz > medián > átlag



Leíró statisztikák: ferdeség (skewness)

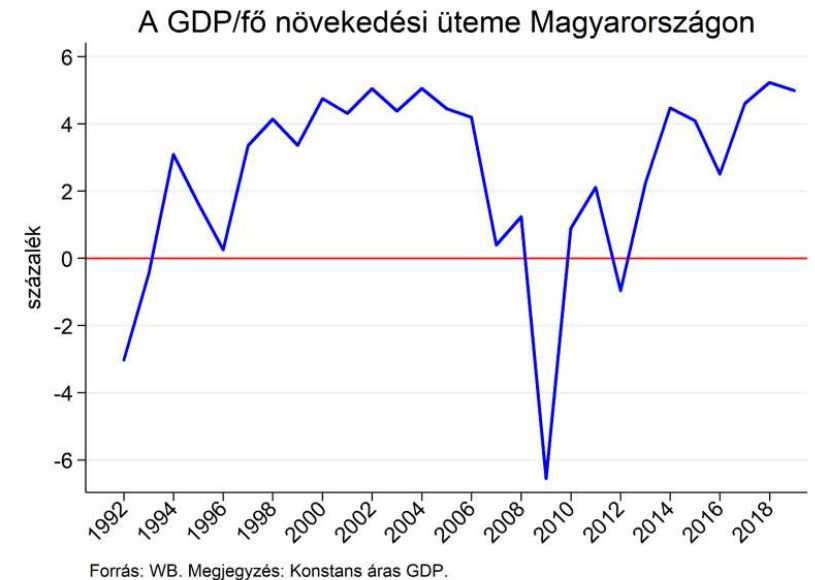


Leíró statisztikák: szóródás



- Terjedelem (range): maximum és minimum közti eltérés
 - Nem megbízható (kiugró értékek) -> interkvartilis (interdecilis) terjedelem = 3. – 1. kvartilis (9. – 1. decilis)

- Szórás
$$s = \sqrt{Var} = \sqrt{\frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N - 1}}$$

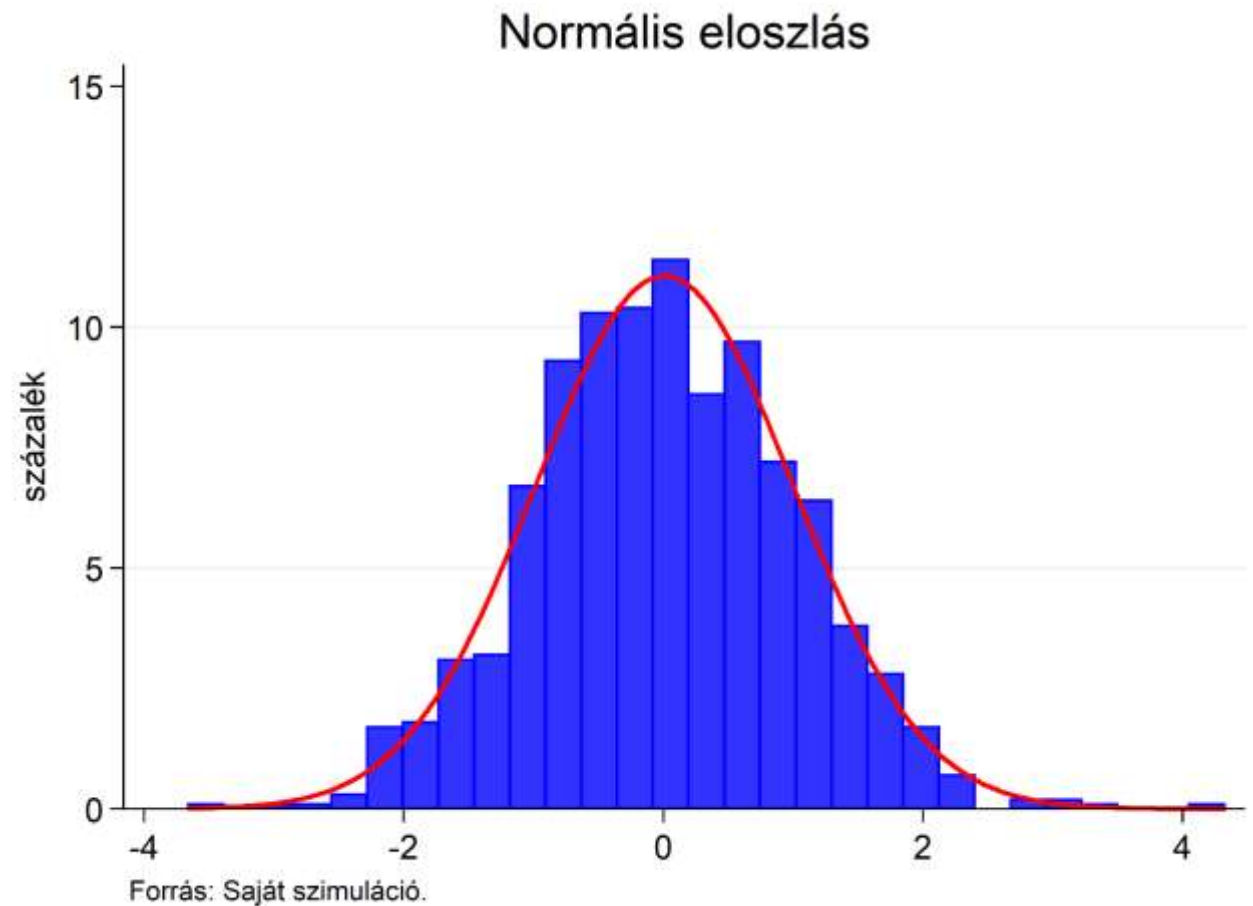


- Variancia: átlagos négyzetes eltérés, azaz szórásnégyzet

A normális eloszlás



- Az adatok
 - 68%-a az $[\bar{Y} - s; \bar{Y} + s]$
 - 95%-a az $[\bar{Y} - 2s; \bar{Y} + 2s]$
 - 99,7%-a az $[\bar{Y} - 3s; \bar{Y} + 3s]$ intervallumba esik.



Leíró statisztikák: szórás – problémák

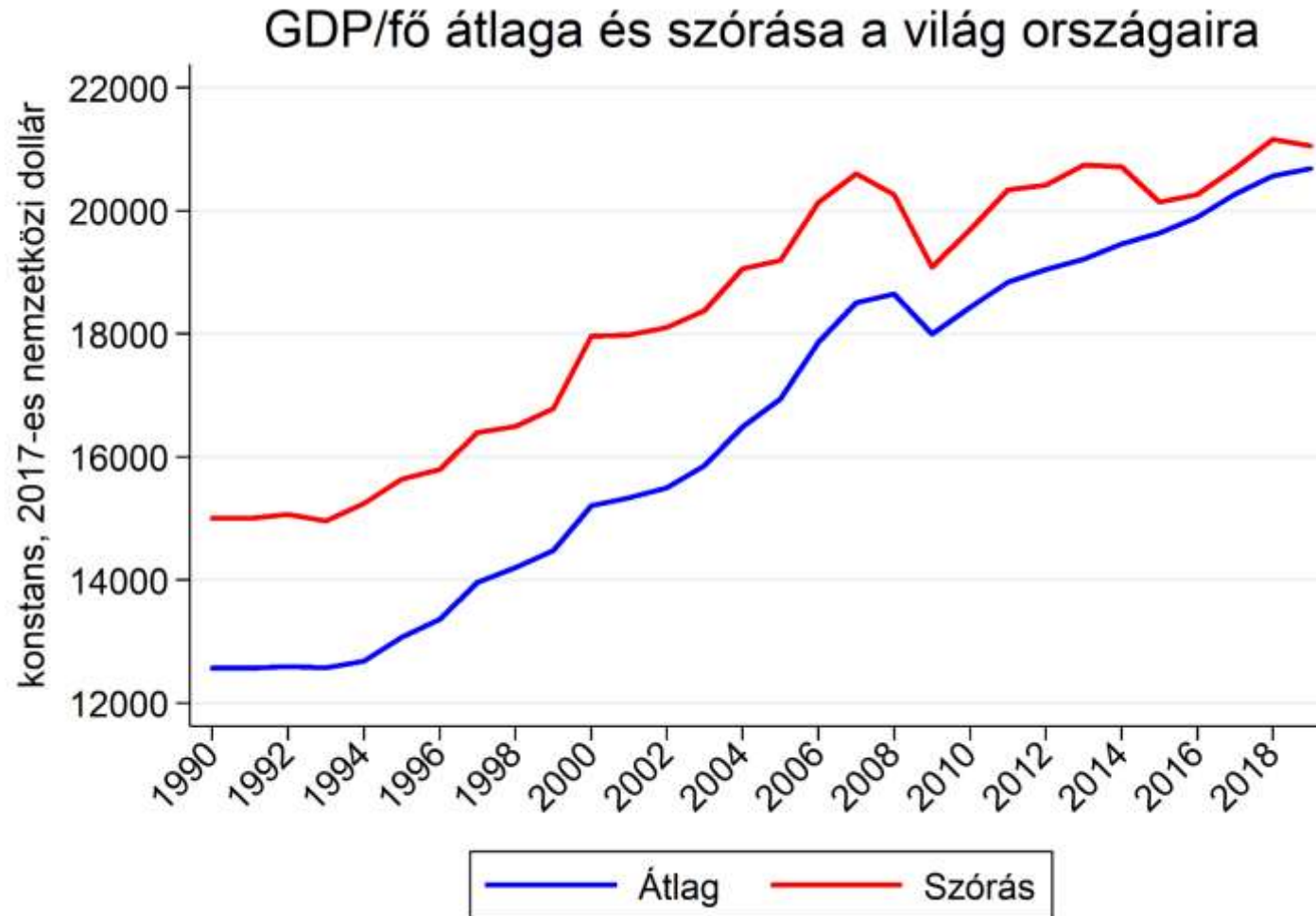


Minek nagyobb a szórása? Az urbanizációnak vagy a demokrácia fokának?

- Urbanizáció: $s=22,7$
- Demokrácia: $s=3,7$

Mértékegysége: ugyanaz, mint az alapadatoké.

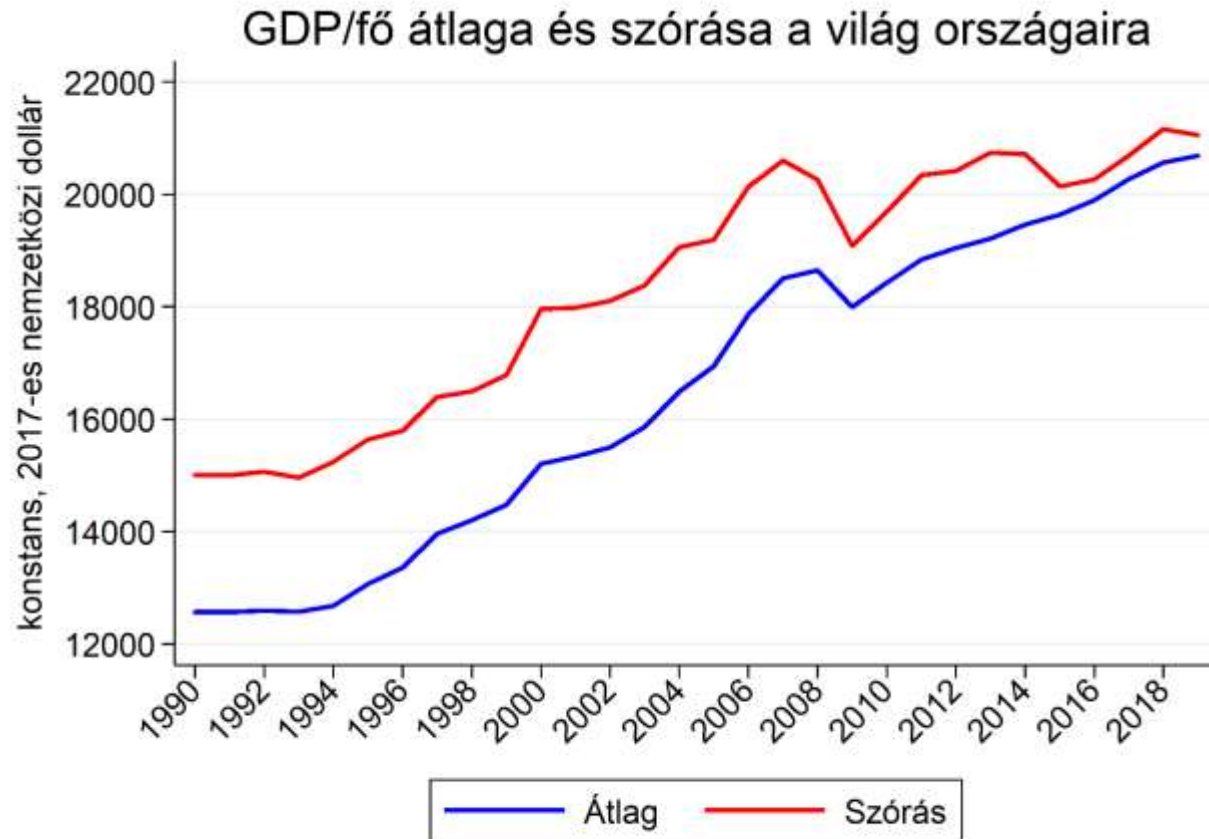
Leíró statisztikák: szórás – problémák



Forrás: Világbank.

2) Csökkent a jövedelem-egyenlőtlenség?

Leíró statisztikák: szórás – problémák



Forrás: Világbank.



Forrás: Világbank.

Relatív szórás: $\text{szórás}/\text{átlag}$

Mértékegysége: %

Feladat



- Nyissa meg az „alkalom_02.xlsx” Excel fájlt, amelyben két munkalapot talál. Mindkét munkalap más adatbázist tartalmaz. Nézze meg az adatbázisokat és hajtsa végre a megadott feladatokat!

Házi feladat



Menjen a Core online-könyvek „Doing Economics”, 1. fejezetéhez: Measuring Climate Change

Készítse el az 1.2 alfejezet feladatait és adja le a megfelelő excel fileokat a Moodle felületén.

Azokon a helyeken, ahol a szöveg ezt kéri (pl. hasonlítsa össze vagy javasoljon, stb.) ott írjon 1-2 mondatot az excel egyik, jól látható cellájába!

<https://www.core-econ.org/doing-economics/book/text/01-02.html#part-12-variation-in-temperature-over-time>

Köszönöm a figyelmet!





Corvinus



Makroökonómia mutatók

Bevezetés az empirikus elemzésbe – 3. hét



Budapesti Corvinus Egyetem

Corvinus University of Budapest

Tartalom



GDP és HDI

Indexek (és infláció)

Koncentrációs mutatók

Munkanélküliség



GDP és HDI



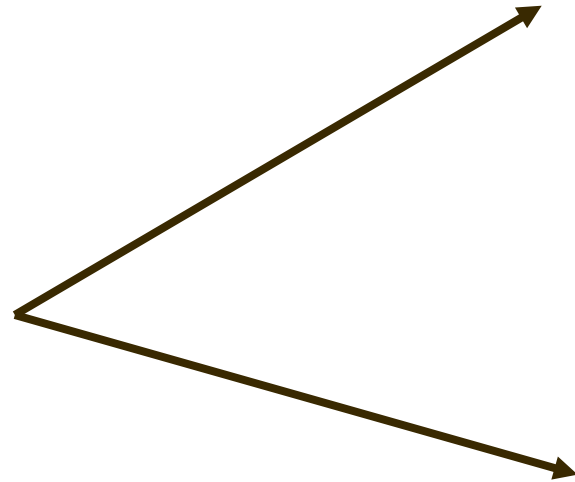
Az aggregált gazdasági teljesítmény meghatározása

A makrogazdasági teljesítmény meghatározása



Értéket számolunk!

A makrogazdasági teljesítmény meghatározása



Végső felhasználást (hozzáadott értéket) vizsgálunk!

A GDP meghatározása termelési oldal

Bruttó hazai termék (GDP): termelési oldalról számítva egyenlő az ágazatok vagy szektorok által előállított, alapáron értékelt bruttó hozzáadott értékek (azaz a kibocsátás és a folyó termelőfelhasználás különbségeinek) és az ágazatokra vagy szektorokra fel nem osztható termékadók és -támogatások egyenlegének összegével.

(A következő slide-okon található információk forrása:
<https://www.ksh.hu/docs/hun/modszgyors/gdpmods15.html>)

A GDP meghatározása termelési oldal (folyt)

Bruttó hozzáadott érték, alapáron:

- + kibocsátás (alapáron)
- folyó termelőfelhasználás (piaci beszerzési áron)
- + termékadókat és -támogatásokat egyenlege

Alapár: az az ár, amennyit a termelő egységnyi termék vagy szolgáltatás értékesítésekor realizál. Azaz az értékesítési árból le kell vonni a termelő által befizetett termékadókat, és hozzá kell adni az értékesítéshez kapcsolódó terméktámogatásokat.

Piaci beszerzési ár: az az ár, amelyet egységnyi termék vagy szolgáltatás igénybevételeért a felhasználó ténylegesen fizet. (Tehát nem tartalmazza a termelő célú felhasználáshoz vásárolt termékek és szolgáltatások után visszaigényelhető áfát, valamint a terméktámogatások értékét.)

A GDP meghatározása termelési oldal (folyt.)

Kibocsátás: az adott gazdasági egység által más, a termelő és szolgáltató gazdasági egységen kívüli egységek számára előállított, valamint a saját végső fogyasztásra kerülő termékek és szolgáltatások összessége. A kibocsátást a nemzeti számlák alapján értékelik.

Folyó termelőfelhasználás: azoknak a termelési folyamat során az elszámolási időszakban más termelőegységtől vásárolt termékeknek és szolgáltatásoknak az értéke, amiket új termékek és szolgáltatások előállításához használnak fel. Az állóeszközök értékcsökkenése azonban nem része a folyó termelőfelhasználásnak. A folyó termelőfelhasználás értékelése piaci beszerzési áron történik.

Bruttó hazai termék (GDP), piaci beszerzési áron

Termelési oldalról:

- + az alapján számított bruttó hozzáadott értékek összege
- + termékadó
- terméktámogatások

Felhasználási oldal

GDP felhasználási oldalról

- + háztartások végső fogyasztási kiadásai
- + államháztartás végső fogyasztási kiadásai
- + nonprofit intézmények végső fogyasztási kiadásai
- + bruttó állóeszköz-felhalmozás
- + készletváltozás
- + export
- import

Modellekben kicsit egyszerűsítve:

$$Y = C + I + G + NX$$

Y : GDP

C : fogyasztás

I : beruházás

G : kormányzati vásárlás

NX : nettó export = EX – IM : export – import

A felhasználási oldal összetevői

Fogyasztás

Háztartások kiadásai termékekre és szolgáltatásokra
Kivétel: ingatlan vásárlása

Beruházás

Termelőeszközök, készletek, építmények vásárlása
Ingatlanvásárlást tartalmazza
Készletfelhalmozás

Kormányzati vásárlások

Helyi, állami vagy szövetségi kormány termékekre és szolgáltatásokra fordított kiadása

Nem tartalmazza a transzfereket

Nettó export = Export - Import

Export: Külföldiek kiadásai belföldön termelt javakra

Import: Külföldi javak vásárlása belföldi lakosok által

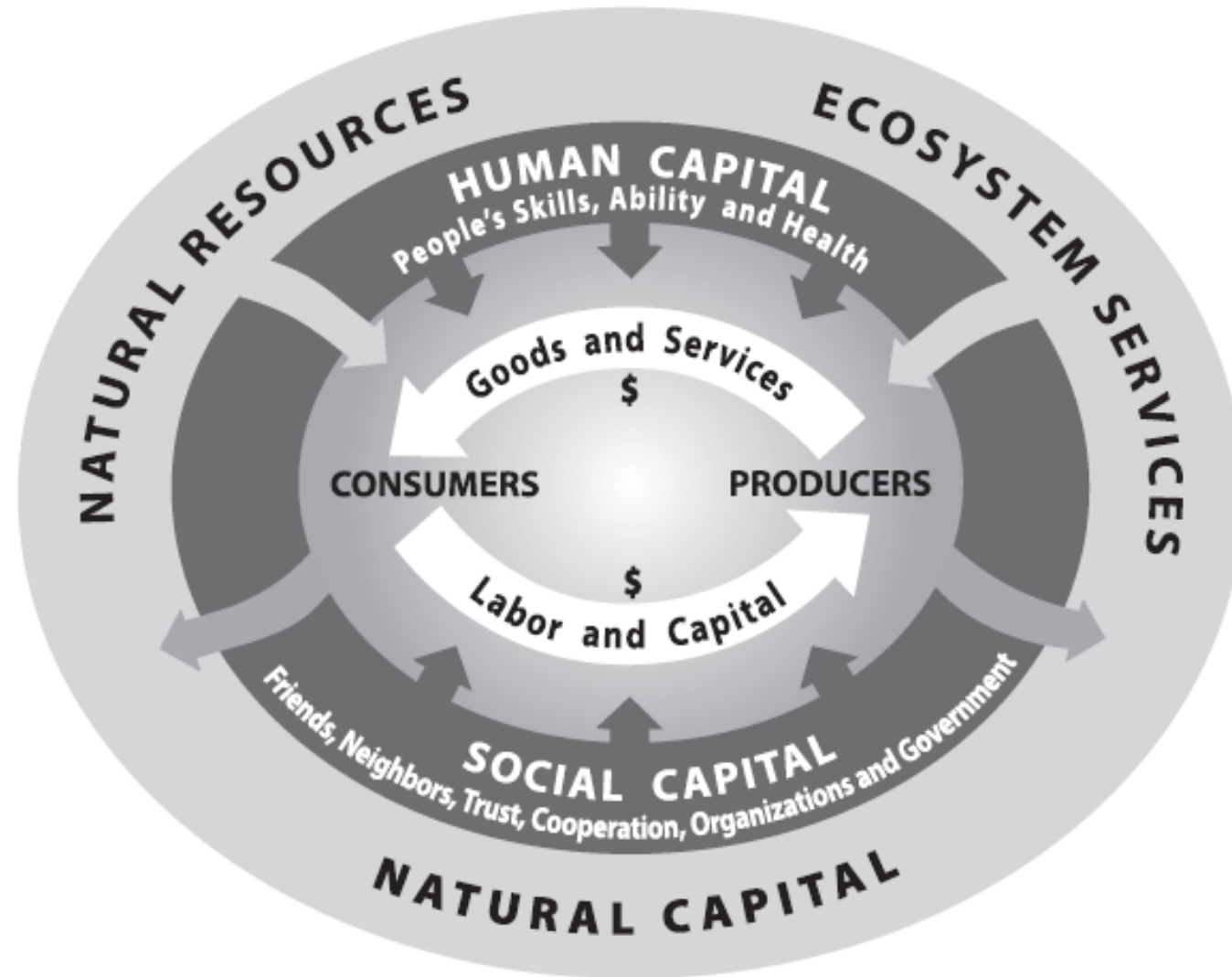
GDP összetevői Magyarországon

2010	Érték (milliárd Ft)	Egy főre jutó (ezer Ft)	GDP százalékában
Háztartások fogyasztása	17236	1724	64,4
Közösségi fogyasztás	2850	285	10,7
Bruttó állóeszköz-felhalmozás („beruházás”)	4806	481	18,0
Készletváltozás	116	12	0,4
Nettó export	1739	174	6,5
Export	23149	2315	86,5
Import	21410	2141	80,0
GDP összesen	26748	2675	100

Jólét mérése

GDP vs. jólét

Figure 2: View of Economy as Part of a Larger System



Jóléti mutatók

- A gazdaság összteljesítménye: GDP.
- Átlagos egyéni jólét gyakran használt mérőszáma: GDP/fő, amely kizárólag gazdasági jólétre koncentrál.
- Human Development Index több aspektusát is megragadja az egyéni jólétnek: Az egy főre jutó jövedelmen túl, figyelembe veszi a várható élettartamot, és az iskolázottságot is.

Élettartamot megragadó index

LE: Várható életartam születéskor (Life Expectancy at birth)

Várható élettartam index:

$$LEI = \frac{LE - 20}{85 - 20}$$

Az index szélső értékei:

$$LEI = \begin{cases} 1, & \text{ha várható életartam születéskor } 85 \\ 0, & \text{ha várható életartam születéskor } 20 \end{cases}$$

<https://ourworldindata.org/life-expectancy>

Iskolázottságot megragadó index

MYS : Azt mutatja meg, hogy a 25 évesek és idősebbek átlagosan hány évet töltöttek tanulással (Mean Years of Schooling).

$$MYSI = \frac{MYS}{15}$$

ahol 15 a 2025-ra adott maximális előrejelzése ennek az értéknek.

EYS : Azt mutatja meg, hogy egy 18 év alatti várhatóan hány évet fog tanulással tölteni (Expected Years of Schooling).

$$EYSI = \frac{EYS}{18}$$

Mivel 18 év a legtöbb országban egy mester diploma megszerzése.

Az oktatási index (Education Index):

$$EI = \frac{MYSI + EYSI}{2}$$

<https://ourworldindata.org/grapher/mean-years-of-schooling-1>

Anyagi jólétet meghatározó index

A GDP az ország területén előálló jövedelem, figyelembe veszi a külföldiek itt keletkezett jövedelmét, és nem veszi figyelembe a magyarok külföldön keletkezett jövedelmét. Ha ezekkel korrigáljuk a GDP, akkor megkapjuk a GNI-t (Bruttó Nemzeti Jövedelem).

A jobb nemzetközi összehasonlíthatósághoz a GNI korrigálható, hiszen az egyes országokban ugyanabból a jövedelemből nem ugyanannyi vásárolható egyes termékekből. Ezt nevezzük a vásárlóerő paritáson számolt jövedelemnek, és az egy főre vetített értékét jelölje: GNI_{pc}

A jövedelem index (Income Index)

$$II = \frac{\ln(GNI_{pc}) - \ln(100)}{\ln(75000) - \ln(100)}$$

Az II értéke 0, ha az $GNI_{pc} = 100\$$, és 1, ha $GNI_{pc} = 75000\$$.

A HDI

Human Development Index képlete ezek után:

$$HDI = \sqrt[3]{LEI \cdot EI \cdot II}$$

<https://ourworldindata.org/human-development-index>

Feladat - 1

1) Nyissa meg az ENSZ statisztikai oldalán a Nemzeti számlákra vonatkozó oldalt:
(<https://unstats.un.org/unsd/snaama/Index>).

2) Töltse le az egy főre jutó GDP-t és összetevőinek adatait 2015-ös fix áron, dollárban, minden országra, minden évre. (Downloads → GDP and its breakdown at constant 2015 prices in US dollar → All countries for all years -sorted alphabetically)

3) Tegyen szűrőt az első sorra, és válasszon ki egy tetszőleges országot!

Mely sorok vonatkoznak a felhasználási oldalra? Számolja ki a nettó exportot, és azonosítsa be a többi komponenst:

GDP= Household consumption expenditure+
General government final consumption expenditure+
Gross capital formation+
(Exports of goods and services – imports of goods and services)

4) A kiválasztott adatokat tegye a „GDP felbontás” nevű munkalapra.

5) Határozza meg minden évben, hogy a GDP 4 komponense mekkora részét adta a GDP-nek (százalékban).

6) Ábrázolja egy ábrán, hogyan alakultak a hozzájárulások az egyes években. Az ábrának legyen címe, legyen látható mi van a tengelyeken, és legyen jelmagyarázat. Az ilyen adatokat hívjuk idősoroknak. Van-e bármilyen mintázat az ábrán, ami logikusnak, vagy éppen meglepőnek tűnik?



Indexek



Indexek

Az index(szám) két aggregátum hányadosa :

$$I_t = \frac{A_t}{A_{t-1}}$$

Ahol A (az aggregált sokaság):

$$A_t = \sum_{i=1}^n q_{i,t} p_{i,t}$$

, ahol q a mennyiséget, p az árat jelöli

Index vonatkozhat árra, mennyiségre és értékre.

Indexek

Tartalmuk alapján:

- Értékindex.
$$I_V = \frac{\sum_{i=1}^n q_{i,t} \cdot p_{i,t}}{\sum_{i=1}^n q_{i,b} \cdot p_{i,b}}$$

- Árindex.
$$I_P = \frac{\sum_{i=1}^n q_{i,t} \cdot p_{i,t}}{\sum_{i=1}^n q_{i,t} \cdot p_{i,b}}$$

- Mennyiségi index.
$$I_Q = \frac{\sum_{i=1}^n q_{i,t} \cdot p_{i,t}}{\sum_{i=1}^n q_{i,b} \cdot p_{i,t}}$$

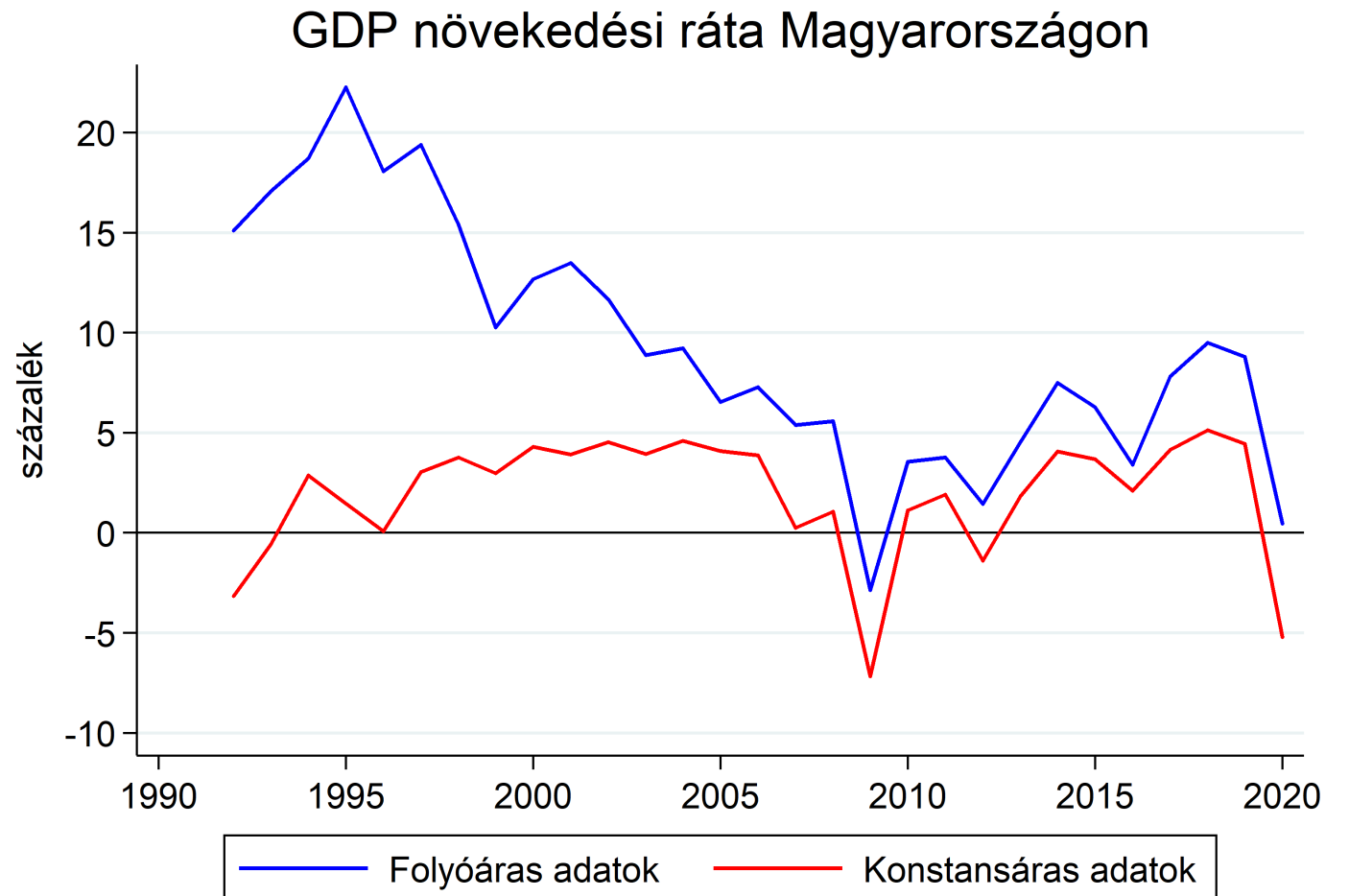
Súlyozás alapján:

- Bázisidőszaki (Laspeyres)
- Tárgyidőszaki (Paasche)

Mennyiségi indexek: egy példa

Árakat rögzítjük, így a mennyiségi változásokat tanulmányozhatjuk.

Reál GDP változása:
$$\frac{\sum_{i=1}^n q_{i,t} p_{i,b}}{\sum_{i=1}^n q_{i,b} p_{i,b}}$$



Forrás: Világbank.



Infláció



Infláció

- Adott ország árainak a változása érdekel. Mit tegyünk? Nem egyes árak alakulása, hanem az országos árszínvonal megfigyelése a célunk.
- Valahogy egybe kell venni a kifli, a fodrász, és a gépkocsi árának a változását. Hogyan? Sima átlagolás nem igazán jó → súlyozott átlag
- Két kérdésre kell válaszolnunk:
 - Milyen súlyokat használunk?
 - Mely javakat vesszük figyelembe?

Árindexek – Milyen súlyokat használjunk?

Több jószágnál a javakat súlyozni kell. A súlyok fontosságot tükröznek.

A súlyok megválasztása alapján két fajta index:

(Ahol b bázisévet jelöl, t pedig tárgyévet)

Laspeyres (bázisidőszaki súlyozás):

$$\frac{\sum_{i=1}^n q_{i,b} p_{i,t}}{\sum_{i=1}^n q_{i,b} p_{i,b}}$$

Paasche (tárgyidőszaki súlyozás):

$$\frac{\sum_{i=1}^n q_{i,t} p_{i,t}}{\sum_{i=1}^n q_{i,t} p_{i,b}}$$

Példa: Árindex

	2009		2010	
termék	p	q	p	q
A	10	0,3	12	0,4
B	4	0,7	5	0,6

Laspeyres árindex

(a tavalyi fogyasztásunk mennyivel lenne drágább idén)

$$\frac{\sum_{i=1}^n q_{i,b} p_{i,t}}{\sum_{i=1}^n q_{i,b} p_{i,b}} = \frac{(0,3 * 12) + (0,7 * 5)}{(0,3 * 10) + (0,7 * 4)} = 1,224$$

Paasche árindex

(az ideai fogyasztásunk mennyivel lett volna olcsóbb tavaly)

$$\frac{\sum_{i=1}^n q_{i,t} p_{i,t}}{\sum_{i=1}^n q_{i,t} p_{i,b}} = \frac{(0,4 * 12) + (0,6 * 5)}{(0,4 * 10) + (0,6 * 4)} = 1,219$$

Feladat - 2

Oldja meg az „Indexek” föl feladatait!

Árindexek – Milyen javakat vegyünk figyelembe?

- Két kérdésre kell válaszolnunk:
 - Milyen súlyokat használunk?
 - Mely javakat vesszük figyelembe?

Képezünk egy tipikusnak tekinthető fogyasztói kosarat, amely egy átlagos fogyasztó által vásárolt termékeket és szolgáltatásokat tartalmazza.

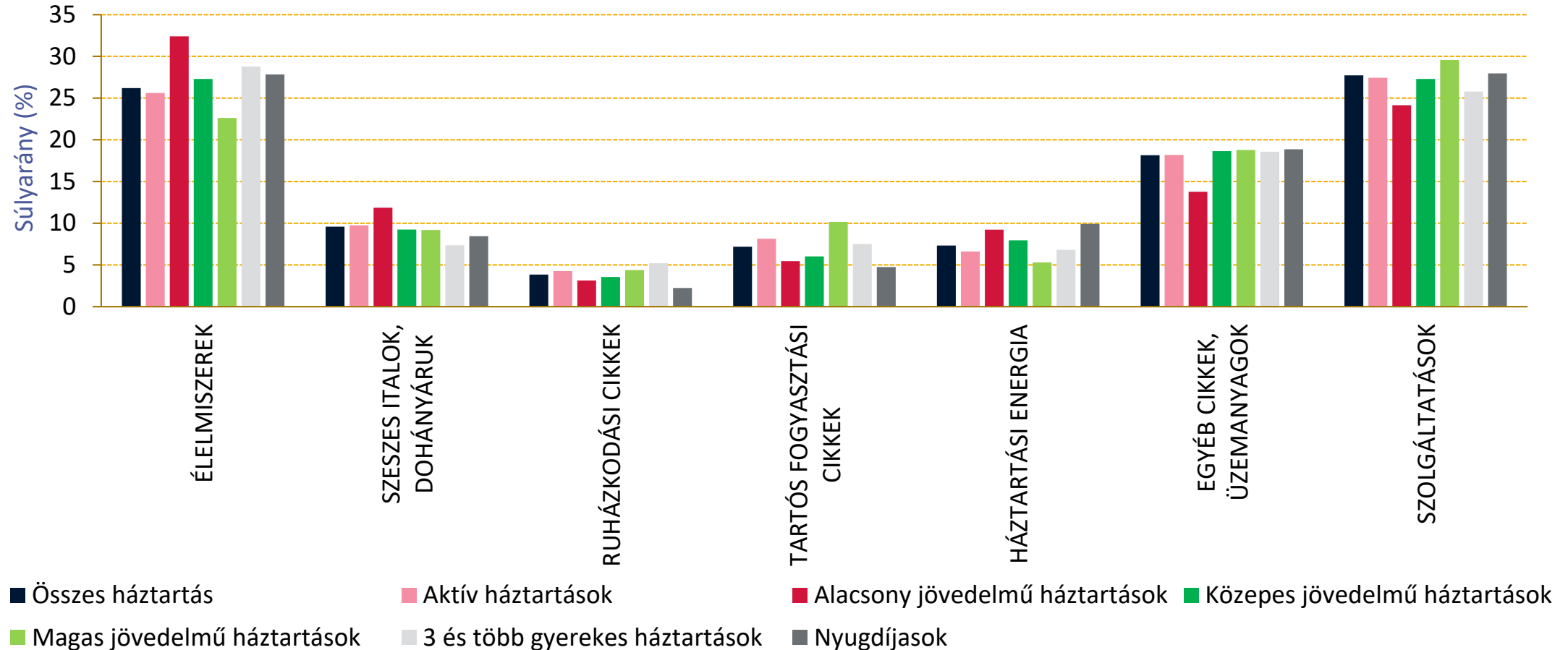
Főcsoportok	A fogyasztóiár-indexben
Élelmiszerek	26,401
Szeszes italok, dohányárúk	10,272
Ruházkodási cikkek	3,795
Tartós fogyasztási cikkek	7,703
Háztartási energia	6,538
Egyéb cikkek, üzemanyagok ^{a)}	17,976
Szolgáltatások	27,315
Összesen	100,000

Forrás: <http://www.ksh.hu/docs/hun/modsz/modsz36.html>

Rétegindexek

Különböző társadalmi csoportok fogyasztóiár-indexei

Fogyasztóiár-index súlyarányok a különböző társadalmi rétegek indexeinél

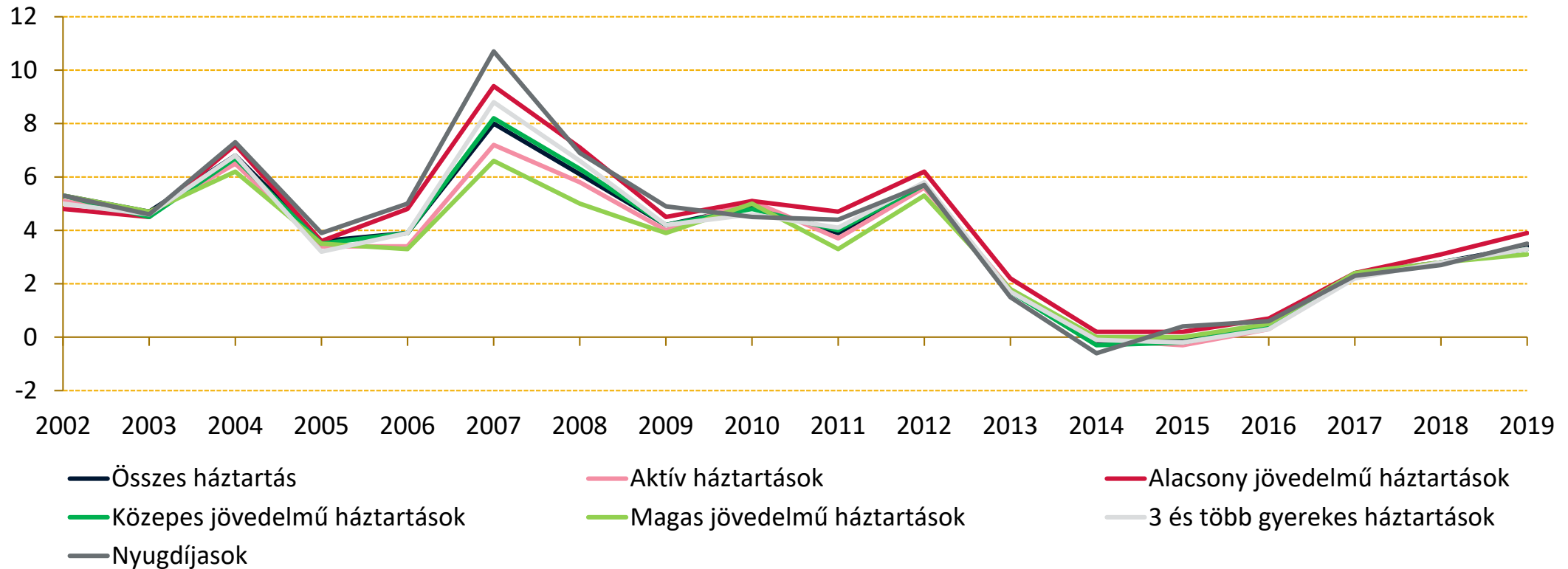


...

Rétegindexek

Különböző társadalmi csoportok fogyasztóiár-indexei

Különböző társadalmi rétegek fogyasztóiár-indexei

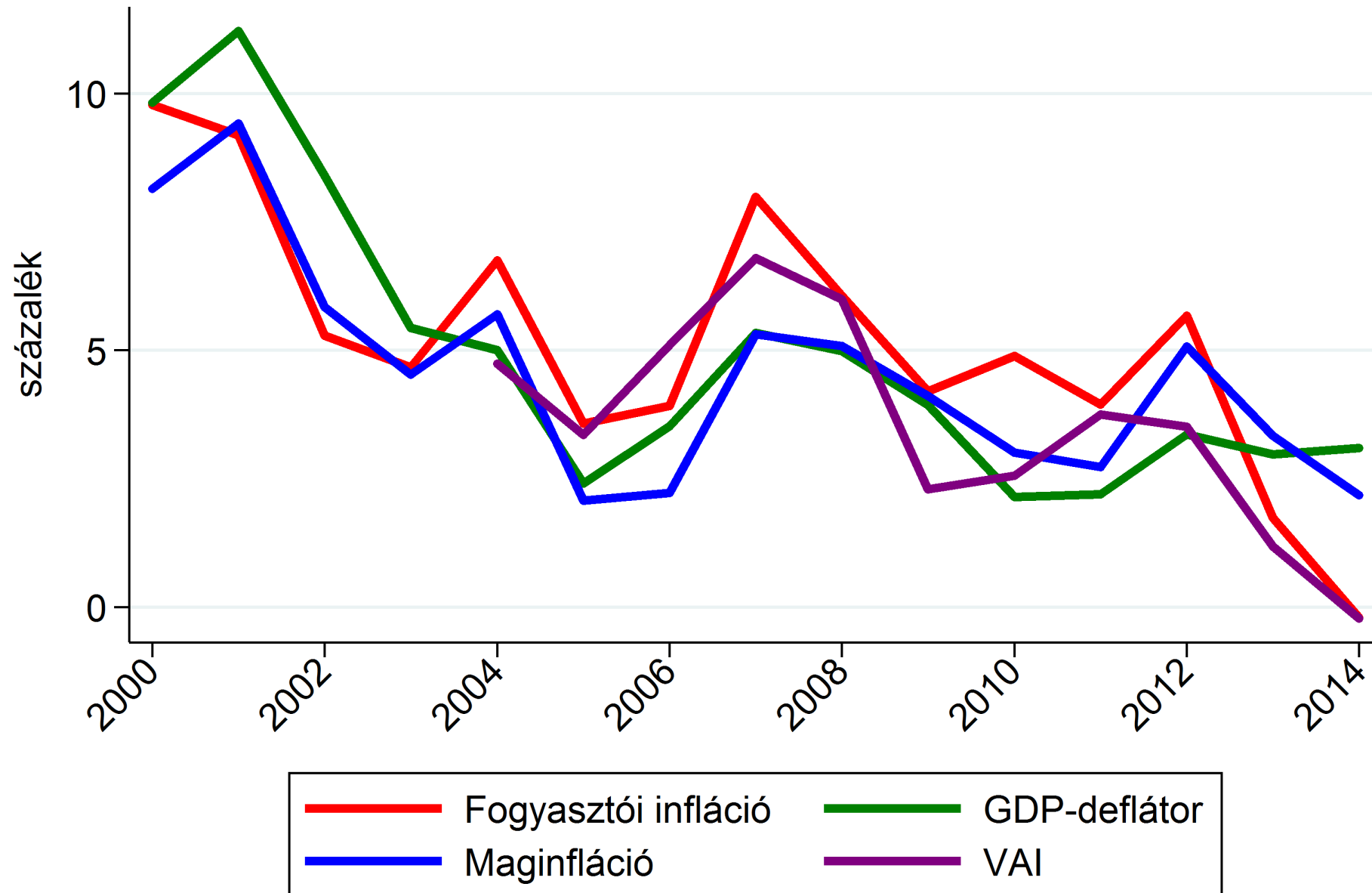


...

Árindexek

- Legfontosabb árindexek: fogyasztói árindex és GDP-deflátor.
- Fogyasztói árindex (consumer price index, CPI):
 - A reprezentatív fogyasztó által fogyasztott jószágkosár árváltozását méri (akár importot is).
 - Bázisidőszaki súlyozású.
- GDP-deflátor:
 - Minden, az adott gazdaságban megtermelt jószágot figyelembe vesz (importot nem).
 - Tárgyidőszaki súlyozású.

Inflációs mutatók Magyarországon



Adatforrás: MNB és World Development Indicators.



Koncentrációs mutatók



Koncentrációs mutatók

Koncentráció: egy adott változó értékének jelentős része vagy egésze kevés egységre összpontosul.

Példák: adott iparágban az összkibocsátás jelentős részét pár vállalat adja; lakossági jövedelem eloszlása egyenlőtlen (80-20 szabály).

Miért fontos? Az idealizált tökéletes versenyben sok, kicsi, árelfogadó vállalat. Valóságban nem tipikus. Sok piacon csak néhány vállalat van, más piacokon jóval több, de néhány meghatározó. Nagy vállalat visszaélhet piaci erejével (nem biztos, hogy tud!) és a fogyasztók rosszul járnak. Versenyhivatalok ezt árgus szemmel figyelik.

Koncentráció – egyszerű mutatók

Állítsunk fel egy csökkenő rangsort a vállalatok között nagyságuk egy bizonyos (minket érdeklő) mérőszáma (pl. piaci részesedés) szerint! Majd kumulálással könnyen kiszámítható, hogy mekkora a piaci részesedése a legnagyobb, az első két legnagyobb, első három legnagyobb stb. vállalatnak.

Ha az előző számolás eredményeit ábrázoljuk, akkor a koncentrációs görbét kapjuk.

Alapötlet: piaci részesedés alapján a teljes kibocsátás mekkora hányada koncentrálódik a legnagyobb vállalatoknál.

CR_n : az n legnagyobb vállalat piaci részesedése. Leggyakoribb a CR_4 .

Példa: pénzügyi intézetek

Magyarországon 2015-ben 40 részvénytársasági formában működő hitelintézet volt. (Adatok: PSZÁF, Aranykönyv 2015). Eszközállomány alapján rangsorolva.

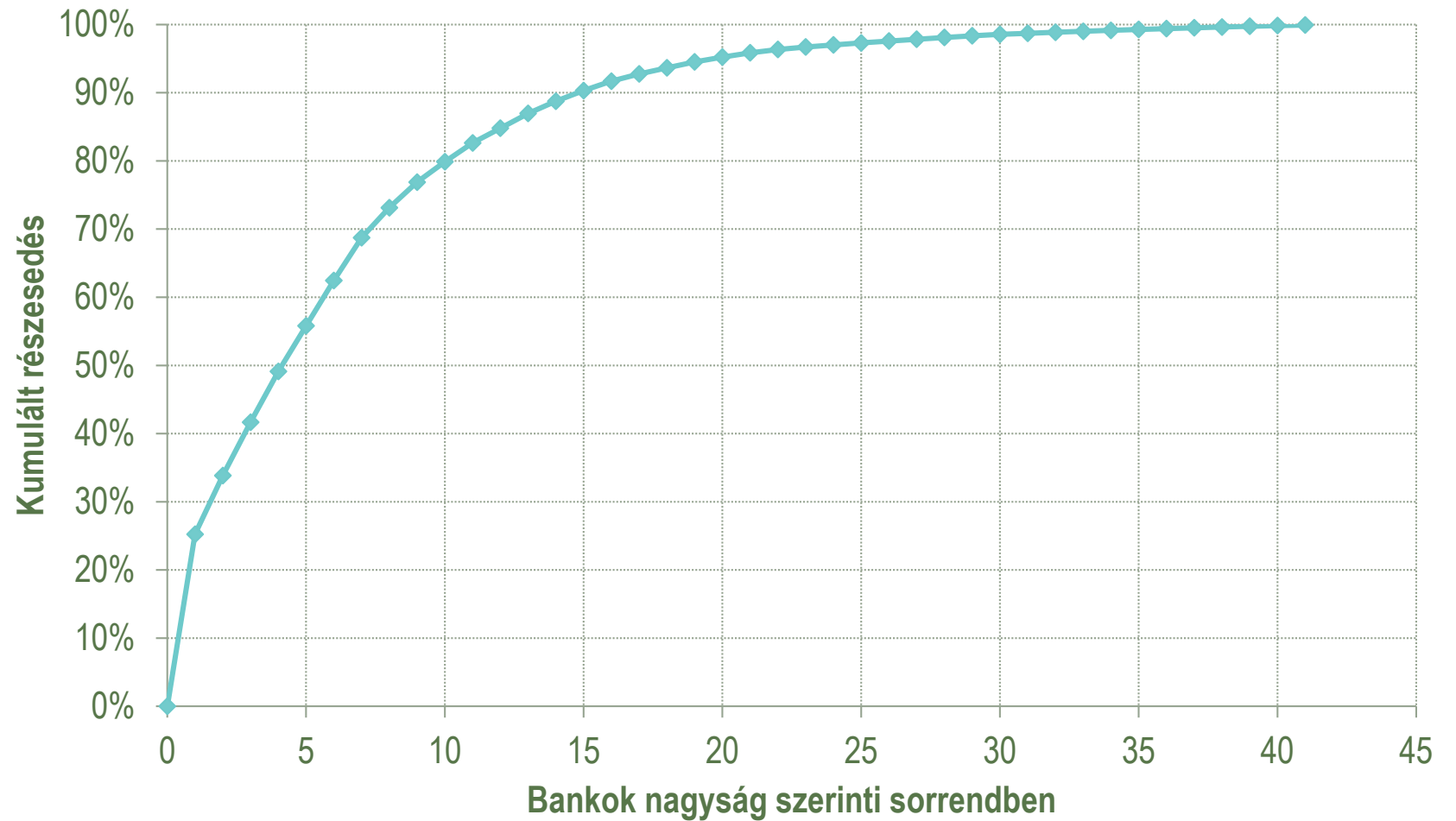
	Név	Piaci részesedés	Kumulált részesedés
1	OTP Bank Nyrt.	25.24%	25.24%
2	K&H	8.60%	33.83%
3	UniCredit Bank Hungary Zrt.	7.83%	41.67%
4	Raiffeisen Bank Zrt.	7.44%	49.11%
5	ERSTE BANK HUNGARY Zrt.	6.68%	55.79%
6	MKB Bank Zrt.	6.67%	62.46%
7	CIB Bank Zrt.	6.28%	68.74%
8	OTP Lakás	4.40%	73.13%
9	MFB Magyar Fejlesztési Bank Zrt.	3.74%	76.88%
10	Budapest Bank	2.99%	79.87%

$$CR_4 = 49.5\%$$

Példa: pénzügyi intézetek

Koncentrációs görbe

	Kumulált részesedés
1	25.24%
2	33.83%
3	41.67%
4	49.11%
5	55.79%
6	62.46%
7	68.74%
8	73.13%
9	76.88%
10	79.87%



Probléma

Előző mutatókkal az a baj, hogy előfordulhat az egyik (CR4) szerint az egyik iparág koncentráltabb, míg a másik (CR8) alapján a másik. Igazából csak a koncentrációs görbe egy pontját ragadják meg ezen egyszerű mutatók.

Olyan mutató kell, ami egy számban fejezi ki a koncentráaltságot, megragadva a koncentrációs görbe egészét.

Ilyen a Herfindahl-Hirschman index vagy egyszerűen Herfindahl-index.

Herfindahl-index

Piaci koncentráció gyakori mérőszáma.

Legyen s_i az i . vállalat részesedése egy olyan piacon, ahol n szereplő van. Ekkor a Herfindahl-index:

$$HI = \sum_{i=1}^n s_i^2$$

Belátható, hogy $\frac{1}{n} < HI < 1$. A határértékek milyen piacszerkezetre utalnak?

Probléma: önmagában a szám nem mond sokat, időbeli alakulása érdekesebb.

Példa: Herfindahl-index

„A” és „B” országban 5-5 autógyár működik a következő részesedésekkel: 10, 14, 16, 28 és 32 %, illetve 12, 13, 14, 30, 31 %. Ha a CR_x mutatókat kiszámoljuk, akkor x függvényében hol az egyik, hol a másik ország autógyártása koncentráltabb.

A Herfindahl-indexek a következők lesznek:

$$HI = 0,1^2 + 0,14^2 + 0,16^2 + 0,28^2 + 0,32^2 = 0,236$$

$$HI = 0,12^2 + 0,13^2 + 0,14^2 + 0,30^2 + 0,31^2 = 0,237$$

(Megj: Néha a 0-10000 intervallumban adják meg a Herfindahl-indexet, azaz $HI_A=2360$ és $HI_B=2370$)

Lorenz Görbe

A koncentráció egyenlőtlenségre utal. Az egyenlőtlenséget gyakran a Lorenz görbe segítségével jelenítjük meg.

Leggyakrabban jövedelmi egyenlőtlenségek bemutatására használják.

A görbe pontjai azt mutatják, hogy az egyének (országok) $x\%$ -a az összjövedelem hány százalékát birtokolja. Minél kevesebb százalék birtokol minél nagyobb hányadot, annál jobban koncentráódik a jövedelem.

(felfelé) kumulálás: egyre több osztályközhöz tartozó értéket adunk össze

Példa: jövedelem

lakosság	jövedelem	kumulált jöv.
1. kvintilis	0,02	0,02
2. kvintilis	0,06	0,08
3. kvintilis	0,13	0,21
4. kvintilis	0,25	0,46
5. kvintilis	0,54	1

Figyelem: a lakosságot rendeztük a jövedelem szempontjából.

Lorenz Görbe

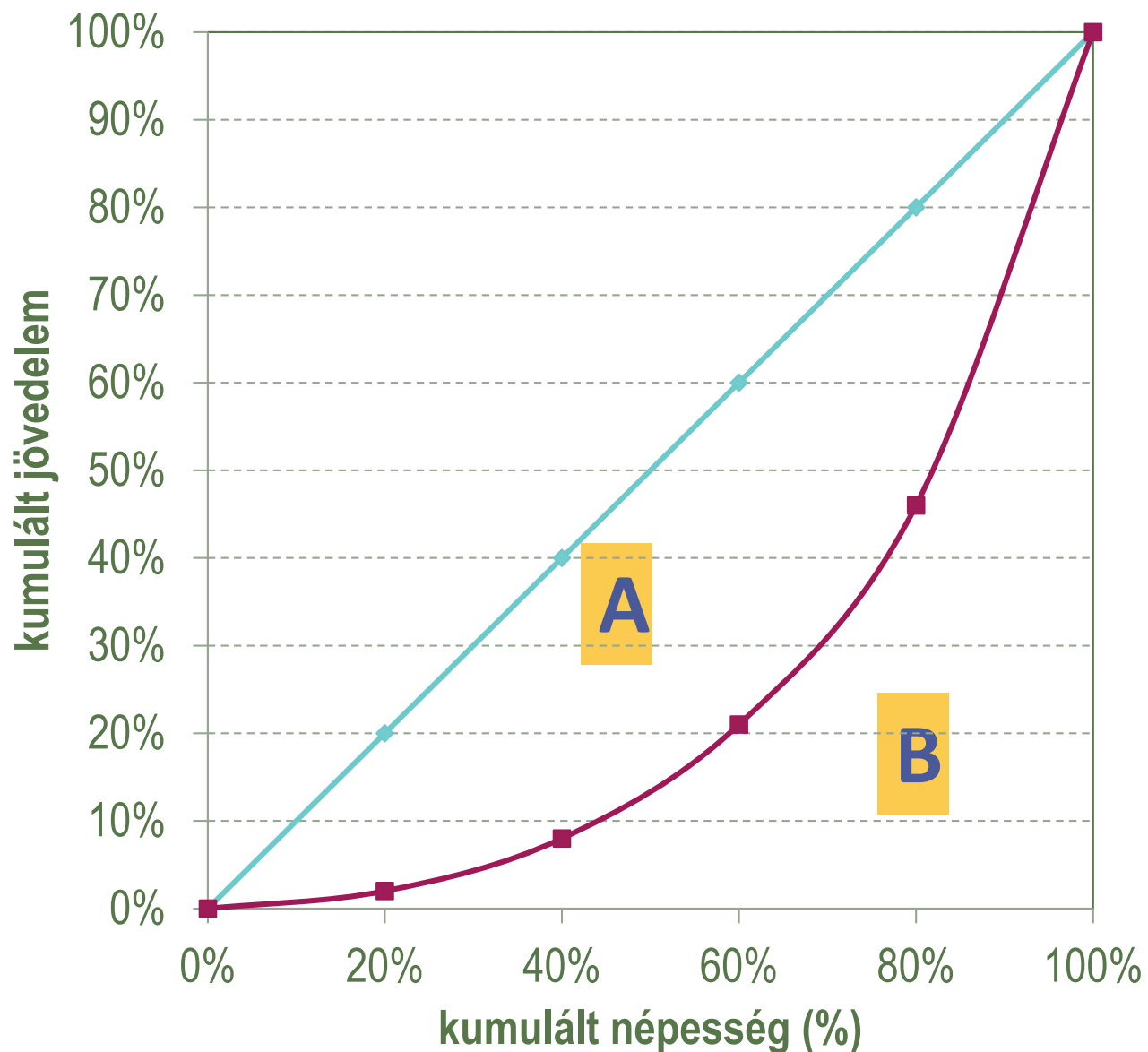
Átló: egyenletes eloszlás

Minél nagyobb az átló és a görbe közötti terület, annál egyenetlenebb az eloszlás (annál jobban koncentrálódik a jövedelem).

Gini együttható: $G=A/(A+B)$

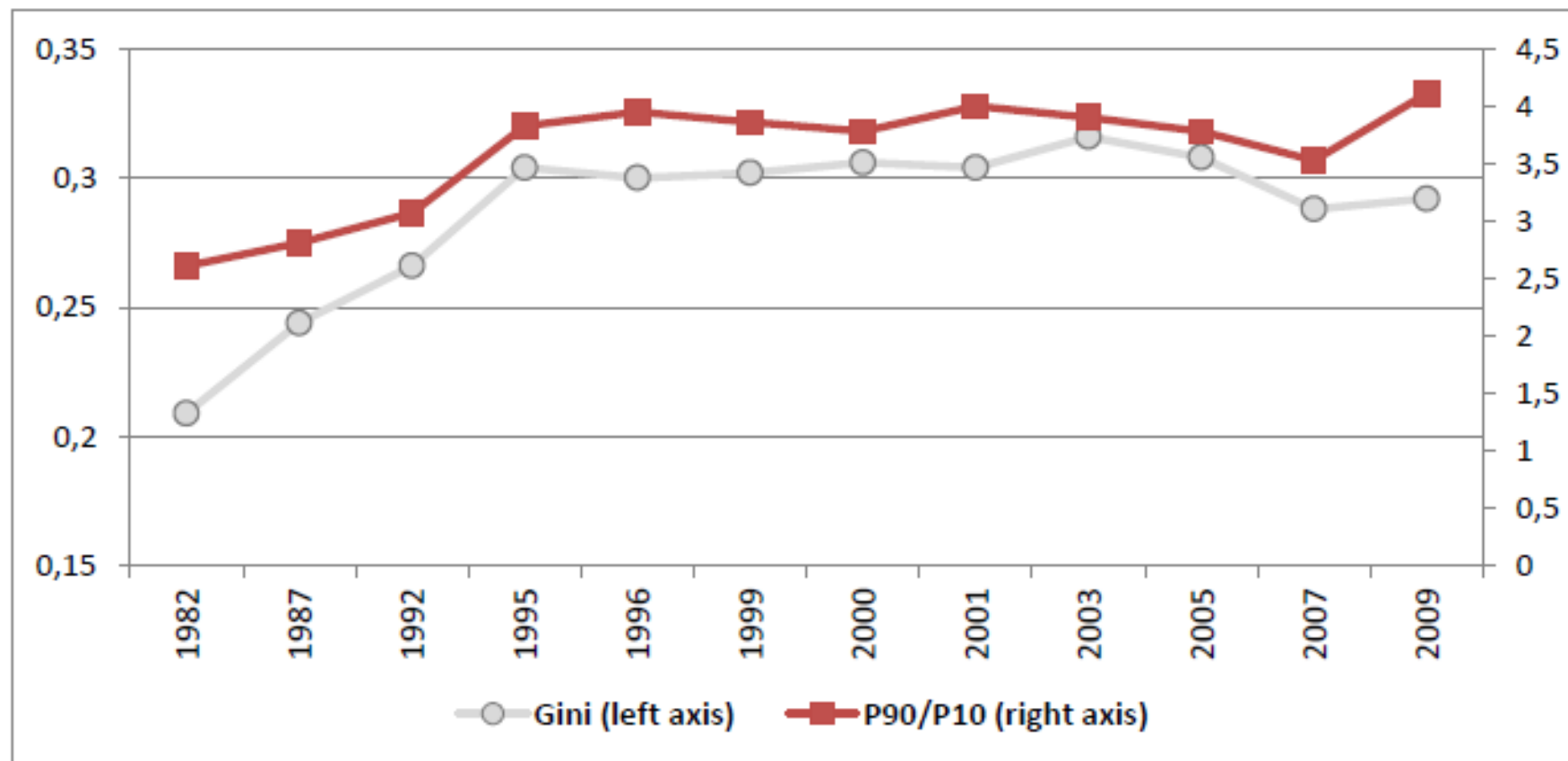
$0 \leq G \leq 1$

Kisebb Gini, egyenletesebb eloszlásra utal.



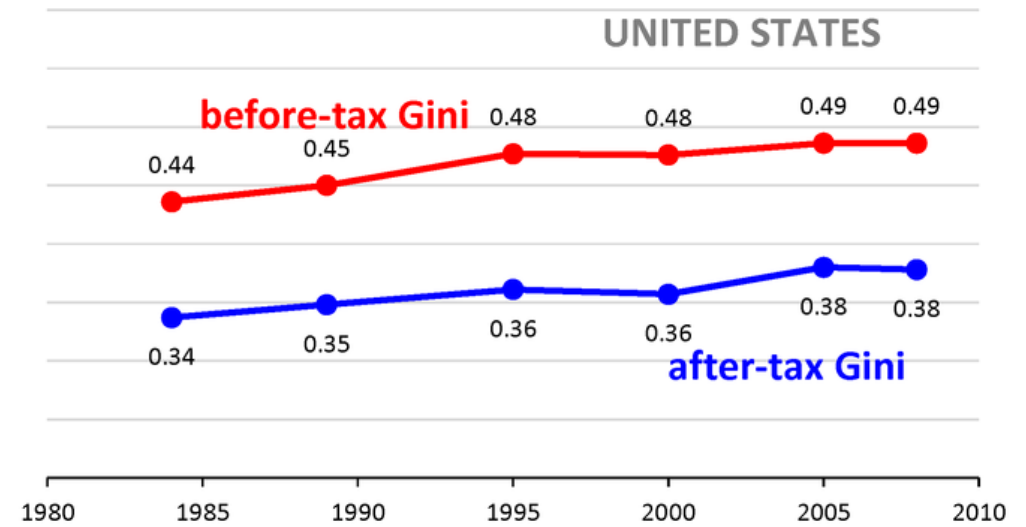
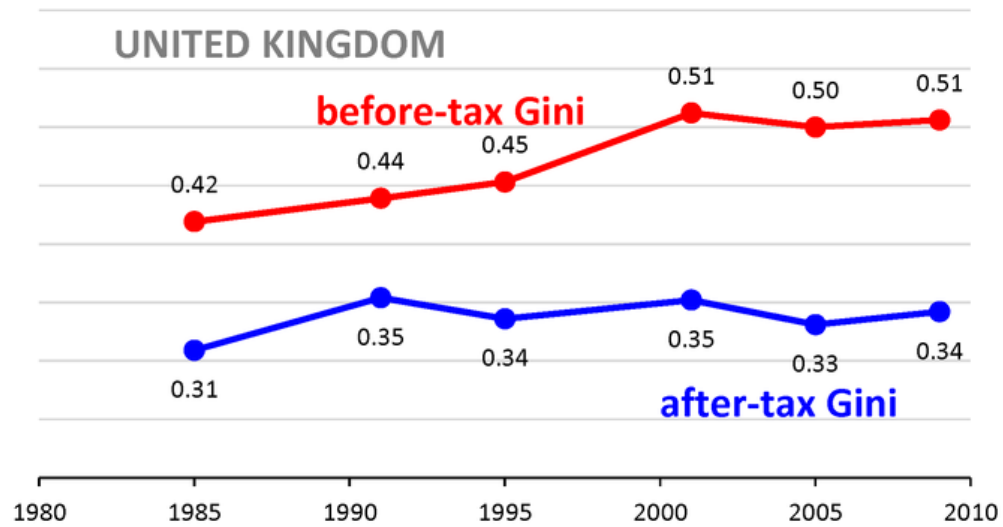
Gini együttható alakulása Magyarországon

Figure 0.1 Long-run evolution of inequality in per capita household income



Source: Tóth (2005, 2010). Data are from: 1962–87: Hungarian Central Statistical Office income survey; 1992, 1995, 1996: Hungarian Household Panel; 1999–2009: TÁRKI Household Monitor.

Gini együttható alakulása adózás előtt és után



Forrás: Wikipedia.

Feladat - 3

Oldja meg az „Koncentrációs mutatók” fül feladatait!



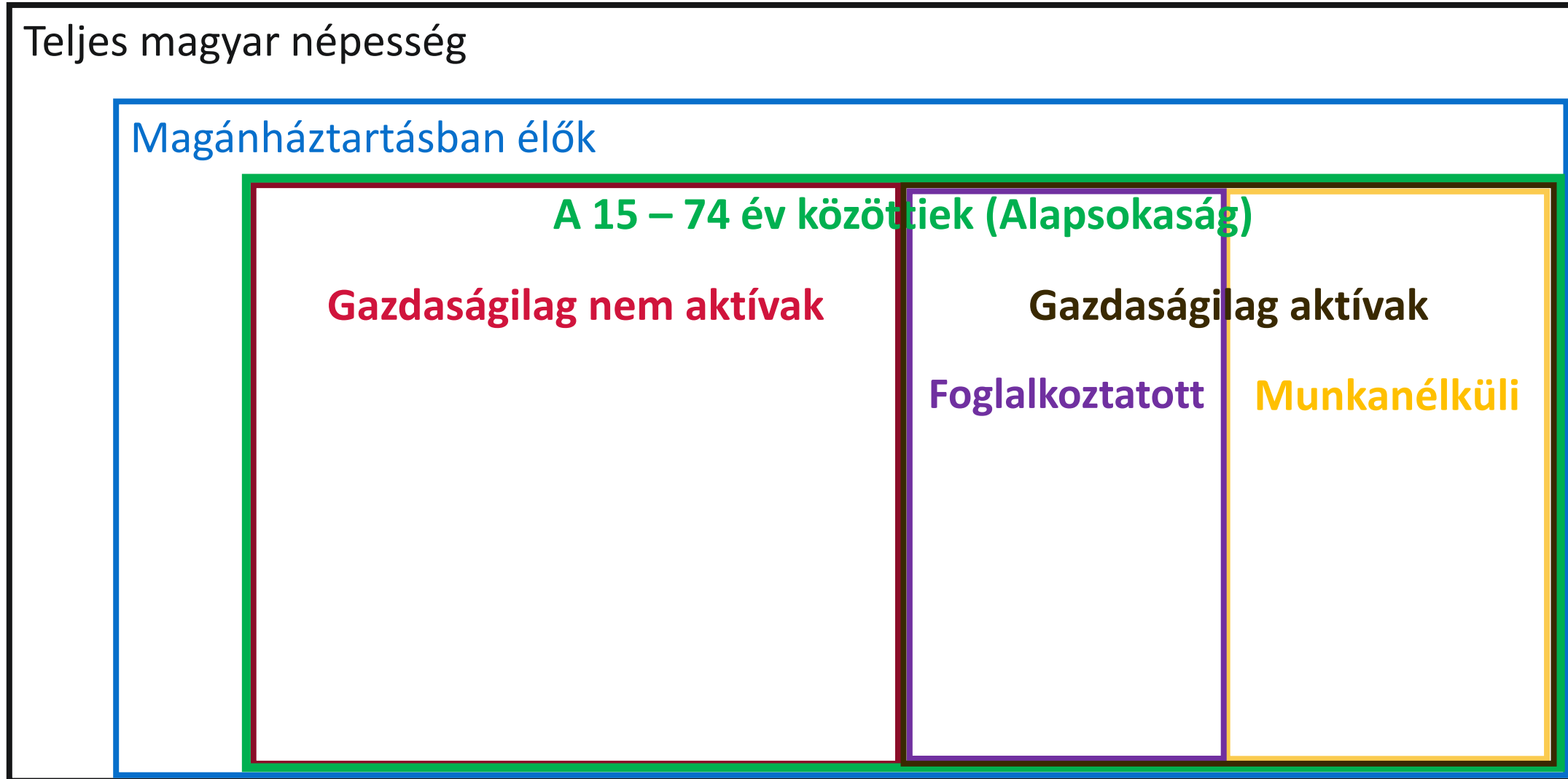
Munkanélküliség



Munkanélküliség

Ki számít munkanélkülinek (statisztikai szempontból), és hogyan azonosítjuk őket?

Összefoglalás



Mintás adatfelvétel

A munkaerő-felmérés (MEF), ami a **magánháztartásokra** kiterjedő reprezentatív felvétel, a **15–74 éves személyek** gazdasági aktivitásáról nyújt információt. Az adatgyűjtés célja, hogy a foglalkoztatottság és a munkanélküliség alakulását a nemzetközi statisztikai ajánlásoknak megfelelően, a **mindenkori munkaügyi szabályozástól, illetve annak változásától függetlenül**, a Nemzetközi Munkaügyi Szervezet (ILO) fogalmait felhasználva figyelje meg.

A magyar munkaerő-felmérés a vizsgált népességet egy meghatározott időszakban (a kikérdezés hetét megelőző héten, a hetet hétfőtől vasárnapig számítva) végzett tevékenységük alapján osztályozza.

Alapfogalmak 1

Foglalkoztatott az, aki az adott héten legalább egy órányi, jövedelmet biztosító munkát végzett, illetve rendelkezett olyan munkahellyel, ahonnan átmenetileg (betegség, szabadság stb. miatt) volt távol.

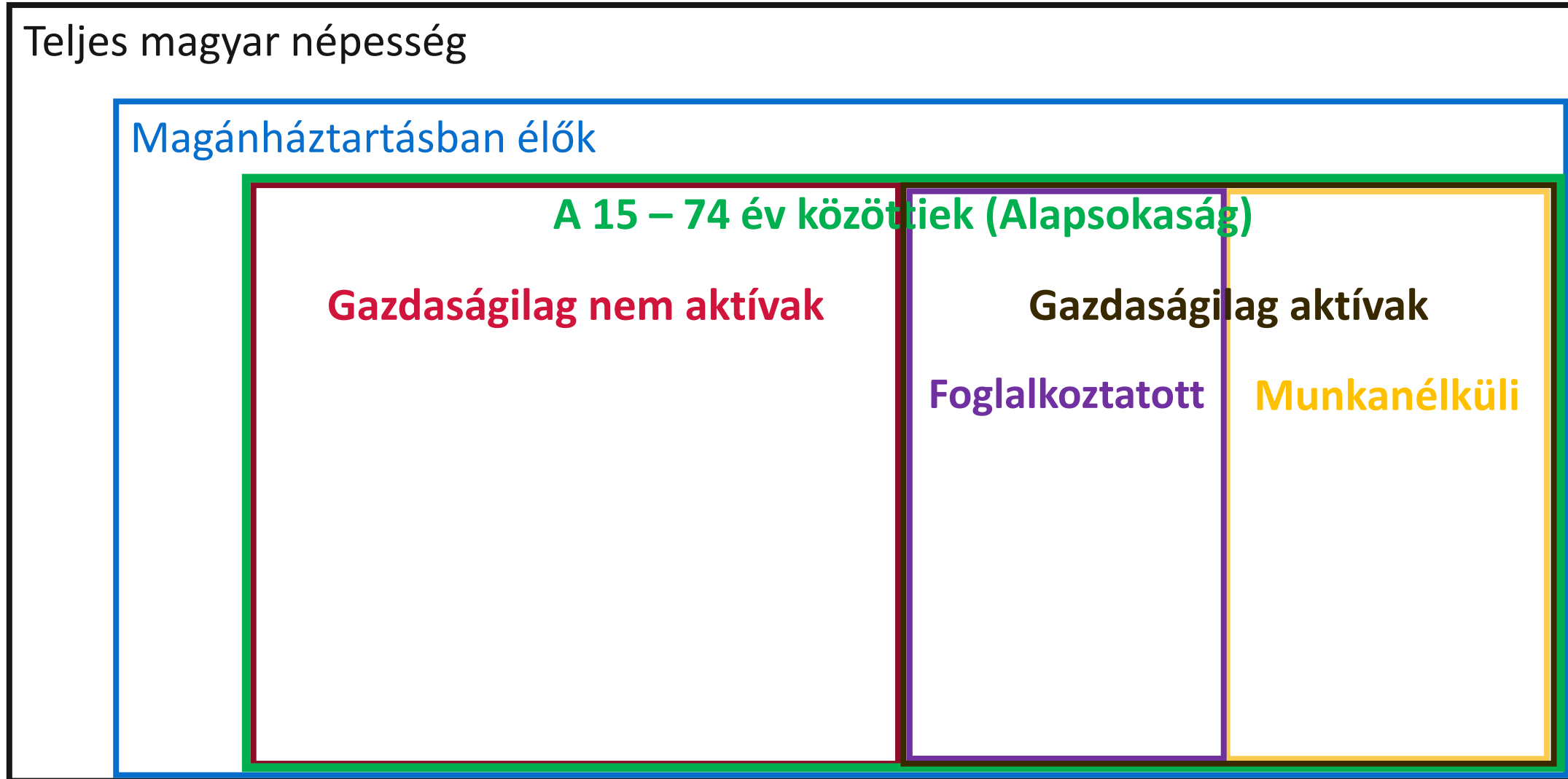
Munkanélküli az, aki az adott héten nem dolgozott, és nincs is olyan munkája, amelyből átmenetileg hiányzott; a kikérdezést megelőző négy hét folyamán aktívan keresett munkát; két héten belül munkába tudott volna állni, ha talált volna megfelelő állást, illetve már talált munkát, ahol 90 napon belül dolgozni kezd.

Alapfogalmak 2

Gazdaságilag aktívak: azok, akik megjelennek a munkaerőpiacon, azaz a foglalkoztatottak és a munkanélküliek.

Gazdaságilag nem aktívak (inaktívak): azok, akik a vonatkozási héten nem dolgoztak, illetve nem volt rendszeres jövedelmet biztosító munkájuk, és nem is kerestek munkát, vagy kerestek, de nem tudtak volna munkába állni.

Összefoglalás



Számított ráták

Aktivitási arány: a gazdaságilag aktívak a megfelelő korcsoportba tartozó népesség százalékában.

Munkanélküliségi ráta: a munkanélküliek a megfelelő korcsoportba tartozó gazdaságilag aktív népesség százalékában.

Foglalkoztatási arány: a foglalkoztatottaknak a megfelelő korcsoportba tartozó népességhez viszonyított aránya.

Feladat - 4

Százholdas pagony lakóiról a következőket tudjuk:

Micimackó: Sajnos 20 évesen már komoly mézfüggőséggel küzd, így jelenleg egy szanatóriumban él és ott kezelteti magát.

Malacka: 19 évesen egy hónapja kezdett dolgozni, egy szanatóriumban, hogy többet lehessen Micimackóval.

Tigris: 22 éves, és épp kiugrott tornatári állásából, de holnap kezd mint cirkuszi akrobata.

Kanga: 28 évesen otthon neveli Zsebibabát, és még 1,5 évig nem is tervez mást.

Zsebibaba: fél éves, és otthon vigyáz, hogy Kanga ne unatkozzon.

Ürge: 50 éves aktív bányász.

Feladat - 4 (folyt)

Róbert Gida: 23 évesen épp befejezte az egyetemet, és alig várja, hogy munkába álljon.

Nyuszi: 30 éves nagytőkés. Dolgozni már csak barátai és üzletfelei szoktak, ő pedig a megtermelt répákat számolgatja.

Bagoly: 40 éves és statisztikát tanít a helyi egyetemen.

Füles: 45 éves, és sajnos egy ideje nem talál munkát. Igazából nem is keres. Ha valaki ajánlani neki egyet arról is azt gondolná, hogy csak átverés.

Számolja ki a százholdas pagony munkanélküliségi rátáját, valamint az aktivitási és foglalkoztatási arányt a „Micimackó” nevű fülön!

Házi feladat

1) Készítse el az alábbi linken található 4.2 fejezet 1,2,3,4,7 feladatait:

<https://www.core-econ.org/doing-economics/book/text/04-02.html?query=GDP#part-42-the-hdi-as-a-measure-of-wellbeing>

2) Készítse el az alábbi linken található 5.1 fejezet feladatait:

<https://www.core-econ.org/doing-economics/book/text/05-02.html>

A feladatok megoldása előtt érdemes elolvasni az adott fejezethez tartó bevezető részeket (Introduction) is. Azokon a helyeken, ahol a szöveg ezt kéri (pl. hasonlítsa össze vagy javasoljon stb.) ott írjon 1-2 mondatot az excel egyik, jól látható cellájába! A feladatokat tagolja, számozza, és derüljön ki, hogy melyik válasz melyik kérdésre vonatkozik!

Köszönöm a figyelmet





Corvinus



Korreláció, egyváltozós regresszió bevezetés

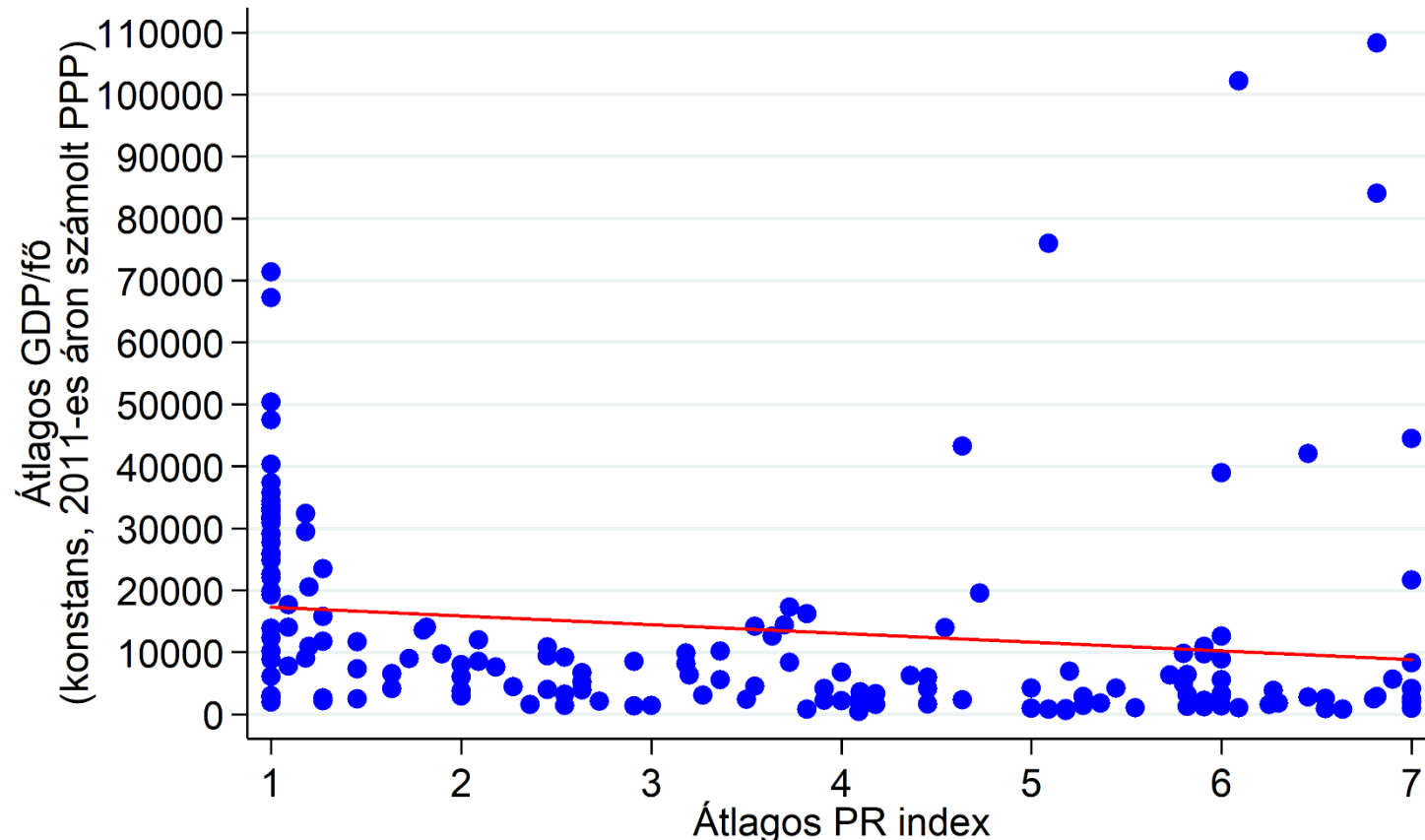
Bevezetés az empirikus elemzésbe – 4. hét



Budapesti Corvinus Egyetem
Corvinus University of Budapest

2. hét vége: Adatok közötti összefüggés bemutatása: pontdiagramm

Átlagos GDP/fő és politikai jogok (PR) index értékek
(1990 - 2000)



- Igazán jó megoldás: számszerű leírás (kapcsolatvizsgálat) – később.

Forrás: Freedom House és Világbank.
Megjegyzés: A magasabb PR érték gyengébb demokráciát jelent.

Tartalom



Korreláció

3

Regresszió

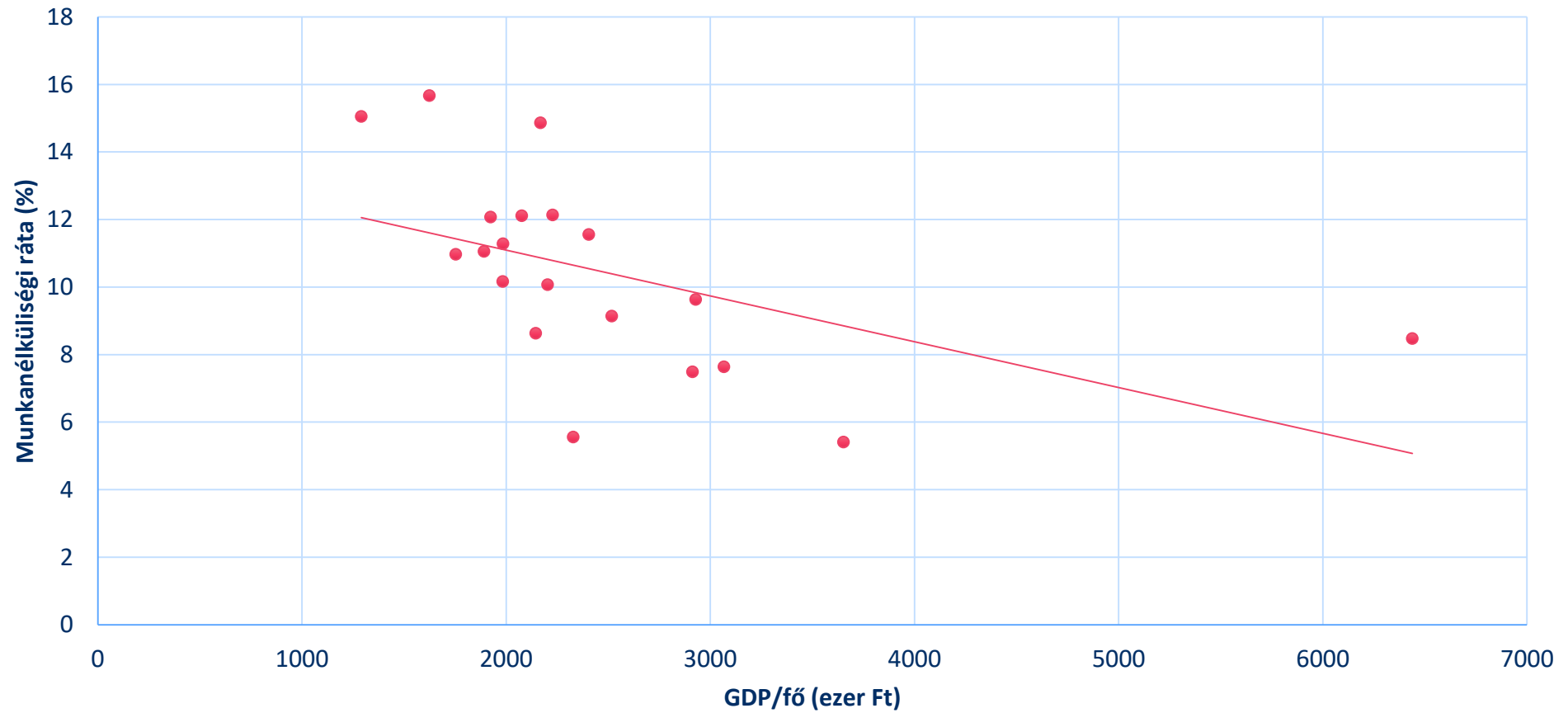
22



Korreláció

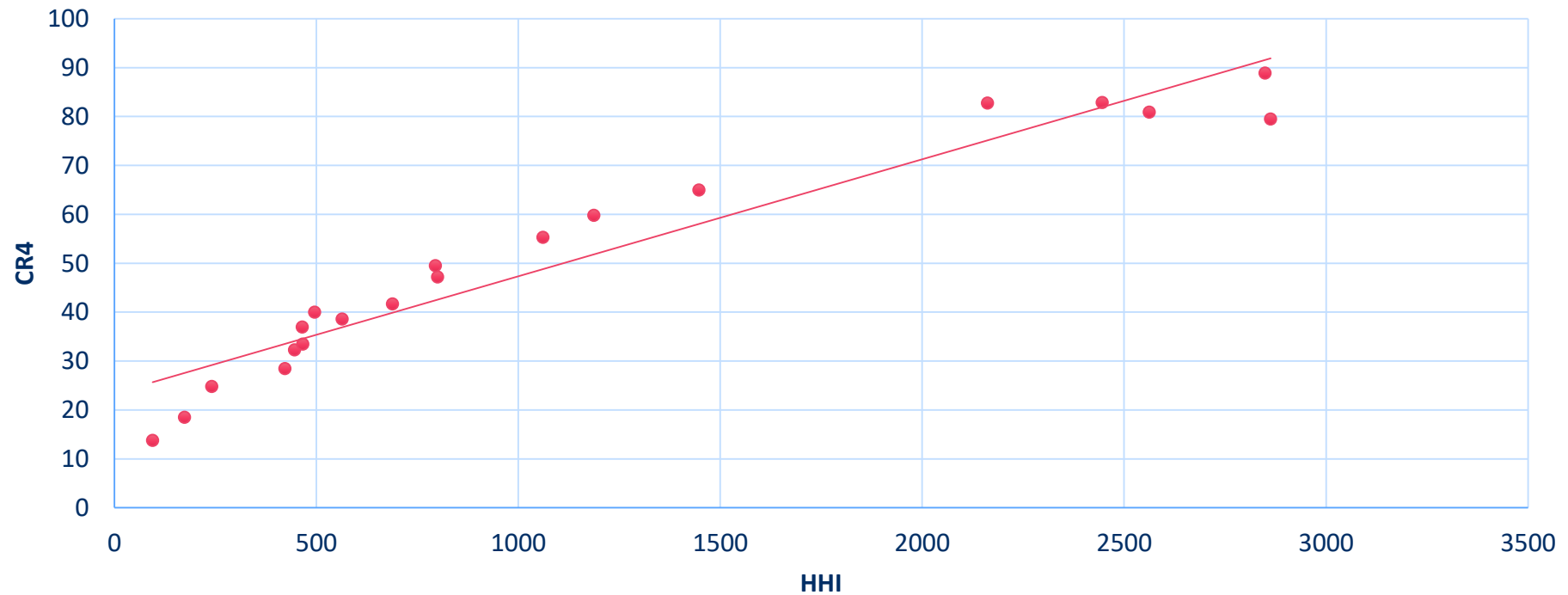


Példa: Két változó között negatív kapcsolat



Példa: Két változó között pozitív kapcsolat

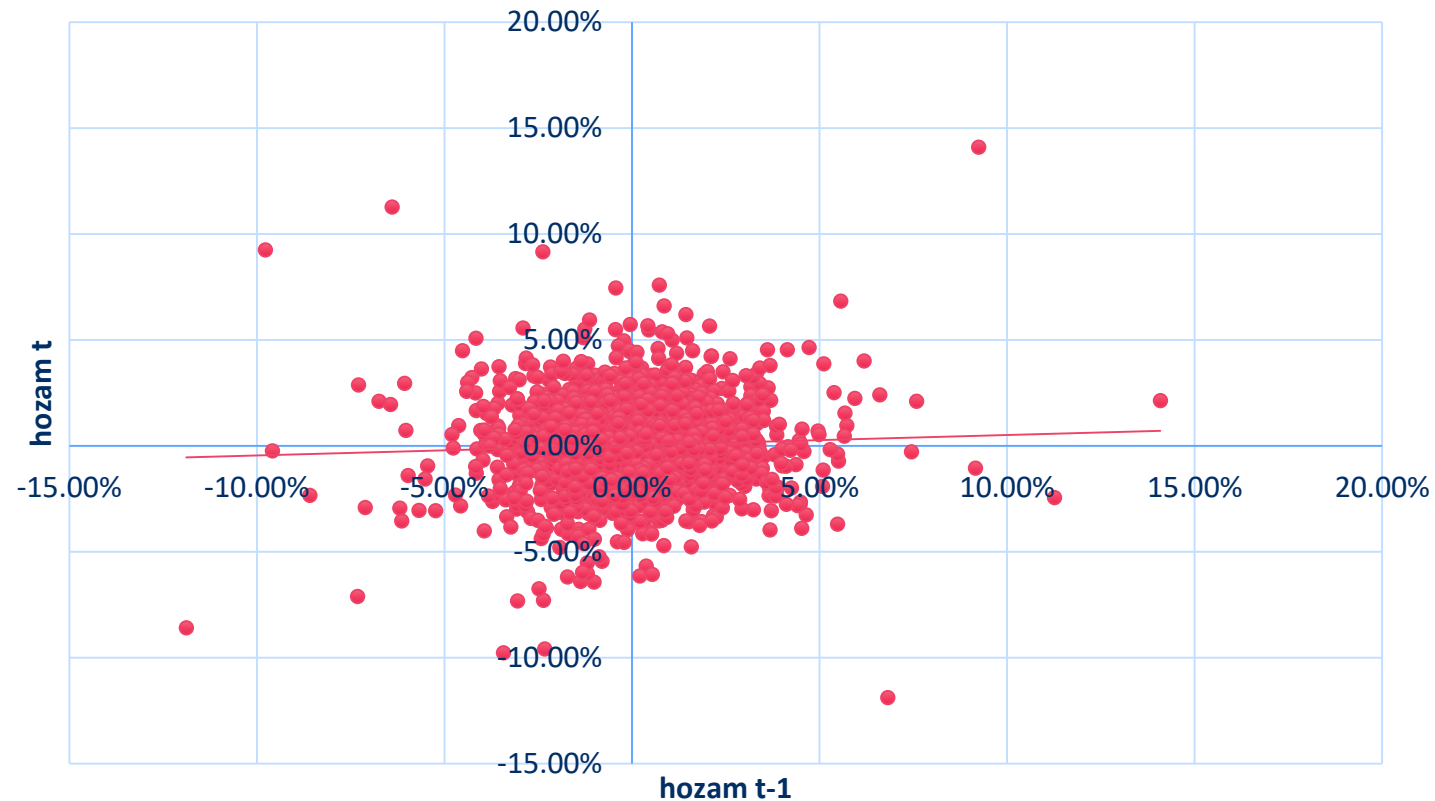
Koncentrációs mutatók néhány iparágban



Példa: Két változó között nincs kapcsolat

BUX egymást követő napi hozamok (2000.09.25 – 2015.09.25)
adott napi hozam előrejelzi-e a következő napit?

Egymást követő napi hozamok összefüggése



Korreláció

Két mennyiségi változó közti kapcsolat *számszerűen*.

Jelölés: X és Y közti korreláció r_{XY}

NEM ok-okozati kapcsolat.

Intuíció: ha az egyik változó magas (alacsony) értékeivel (mihez viszonyítva?) a másik változó magas (alacsony) értékei „járnak együtt”, akkor pozitív a korreláció. Ha az együttjárás ellentétes, akkor negatív. Ha pedig nem mozognak együtt, akkor 0 (nincs korreláció).

A lineáris korrelációs együttható

$$r = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})}{\sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2} \sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}}$$

- A kovariancia a szórások négyzetének gyökével normálva.

$$r = \frac{C}{S_X \cdot S_Y}, \text{ ahol}$$

$$C = \frac{\sum_{i=1}^n (Y_i - \bar{Y}) \cdot (X_i - \bar{X})}{n}$$

$$S_X = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (X_i - \bar{X})^2}$$

Korreláció tulajdonságai

- $-1 \leq r_{XY} \leq 1$
- Pozitív (negatív) érték jelentése: pozitív (negatív) kapcsolat. $r=0$: nem korrelál a két változó.
- Nagyobb érték jelentése: erősebb kapcsolat
- X és Y közti korreláció = Y és X közti korreláció
- Változó korrelációja önmagával = 1
- Konstanssal korreláció = 0
- Korreláció négyzete (r_{XY}^2):
 - Y varianciájának mekkora hányadát magyarázza X varianciája
 - X varianciájának mekkora hányadát magyarázza Y varianciája
 - Az X változó egyes egységénél előforduló X_i értékeinek ismerete hány százalékkal csökkenti az Y változó Y_i értékeinek becslésekor elkövetett hibát (ha ez a becslés lineáris regresszió segítségével történik).

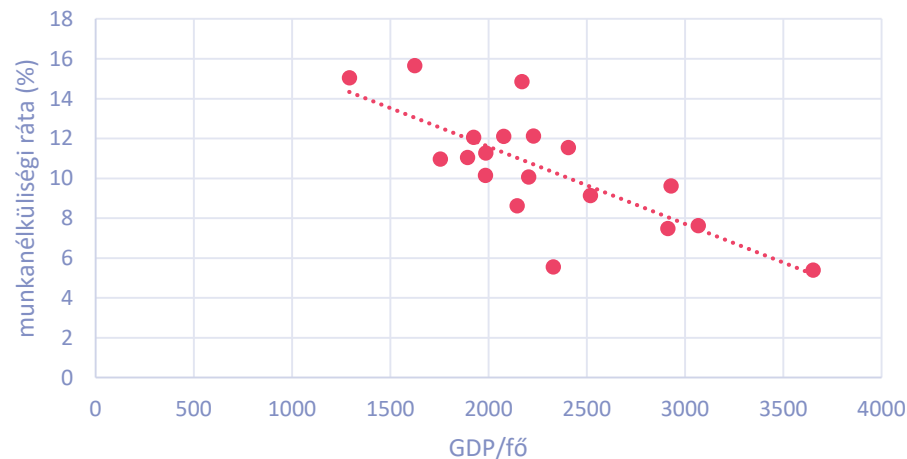
Példa

- Munkanélküliségi ráta és GDP/fő közti korreláció = -0,51
- Általános tendencia, lehetnek eltérések.
- Negatív kapcsolat
- Magasabb GDP/fő – alacsonyabb munkanélküliség
- GDP/fő megyénkénti variáciája a munkanélküliségi ráta variáciájának 26%-át magyarázza (vagy fordítva) – ezt a korreláció négyzete (r^2) fejezi ki ($0,51 \cdot 0,51 = 0,26$)

GDP/fő és munkanélküliség magyar megyékben
(2. óra órai feladata)



1. feladat



Outlierek szerepe!

- Budapest nélkül: $r = -0,75$
- Budapest nélkül: $r^2 = 0,56$

Két példa

Hirdetés és eladás közötti összefüggés

Házárak és telekméret közötti összefüggés

Adatok Gary Koop könyvéhez tartozó adatokból:

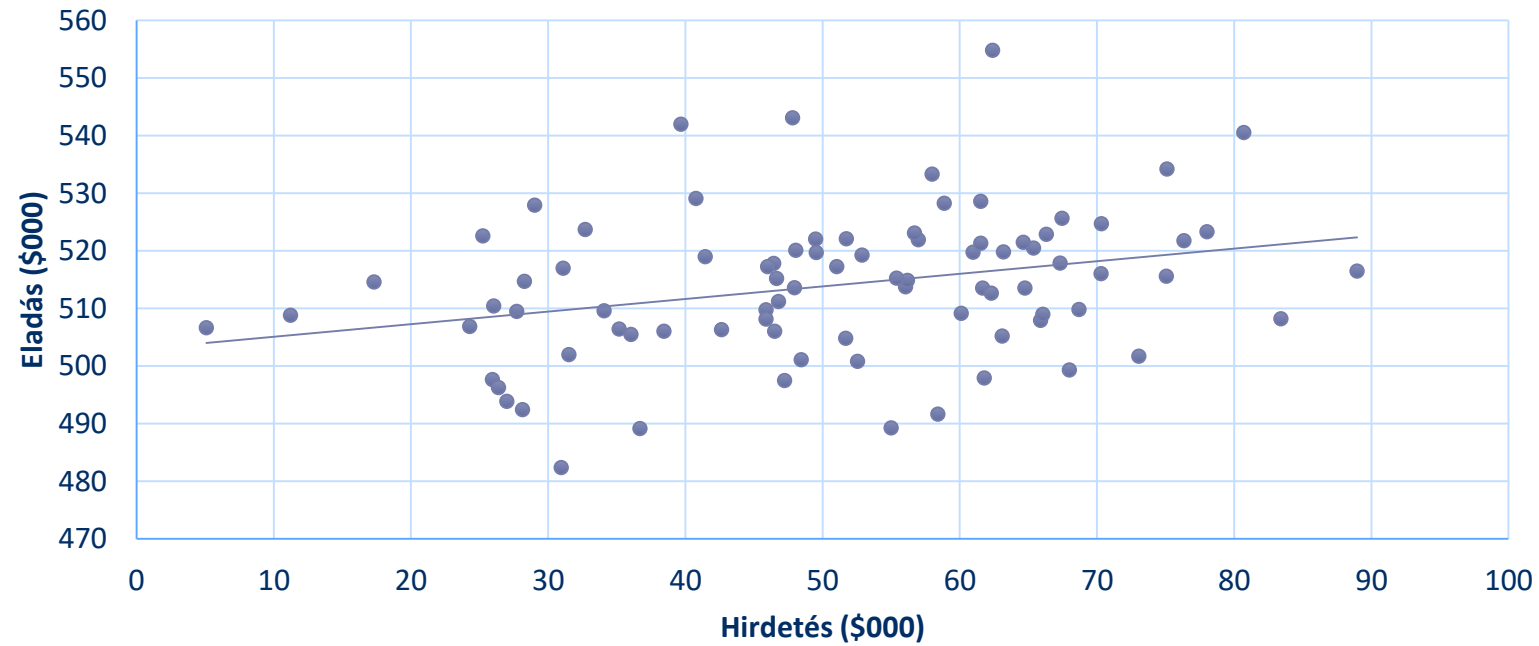
<http://eu.wiley.com/legacy/wileychi/koopdata2ed/dataset.html>

Letölthetőek ingyen. Fogunk velük dolgozni.

Hirdetés és eladás

Keresztmetszeti minta vállalatokról (N=84)

Hirdetések és eladások kapcsolata



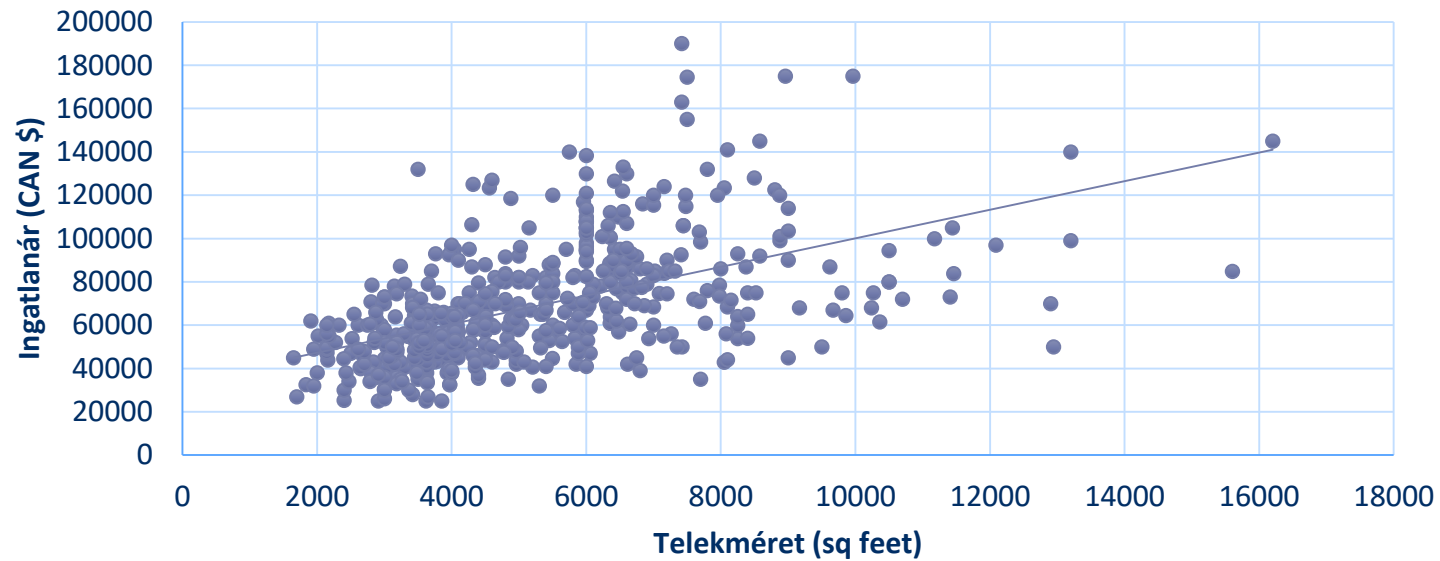
Korreláció=0.298, $r^2=0.089$ – értelmezés?

Mi lehet a mögöttes sztori?

Házárak és telekméret

Keresztmetszeti minta a kanadai Windsorból (N=546, 1987)

Ingatlanárak és telekméret kapcsolata Windsorban



Korreláció=0.536, $r^2=0.287$ – értelmezés?

Mi lehet a mögöttes sztori?

Feladat - 1

Oldja meg a „korreláció” fülön található feladatokat!

Okság

- Egyik változó „okozza-e” a másikat közvetlenül?
- Korreláció önmagában nem árulkodik az okság meglétéről és irányáról.
- Példa: telek nagysága vs. ingatlanár?
- Példa: jövedelem vs. iskolai végzettség?
- Példa: reklámköltség vs. értékesítésből származó bevétel?

- Van okság? Ha van, milyen irányú? Esetleg harmadik mögöttes ok?

Harmadik mögöttes ok

Tk.: dohányzás, alkoholfogyasztás és tüdőrák kapcsolata. Egyértelműnek tűnik, hogy a dohányzás (és még sok minden más) okozza a tüdőrákot. Kettő között a korreláció pozitív. Sokan akkor dohányoznak, amikor isznak is (kocsma), így dohányzás és alkoholfogyasztás is pozitívan korrelál. Ezért minden bizonnyal pozitív a korreláció az alkoholfogyasztás és a tüdőrák között, de biztosan nem az alkohol okozza a tüdőrákot.

Gólyák száma és született gyerekek száma között pozitív a korreláció. Gólya hozza a gyereket?

A szappanok száma és az ingatlanárak között pozitív a korreláció (valószínűleg). Miért lehet?

Több változó közti korreláció

M db változó \rightarrow $M(M-1)/2$ db korreláció

Korrelációs mátrix 3 változóra (X, Y, Z):

	X	Y	Z
X	1		
Y	r_{XY}	1	
Z	r_{XZ}	r_{ZY}	1

Több változó közti korreláció

Ingatlanárak, telekméret és hálósobák közötti korrelációs mátrix:

	<i>Ingatlan ár</i>	<i>telekméret</i>	<i>#hálószoba</i>
<i>Ingatlan ár</i>	1		
<i>telekméret</i>	0.535796	1	
<i>#hálószoba</i>	0.366447	0.151851	1

Több változó közti korreláció

Bértarifa adatok (2011): 45652 megfigyelés alapján

	férfi	kereset	tapasztalat	egyetem
férfi	1.000			
kereset	0.086	1.000		
tapasztalat	-0.044	-0.067	1.000	
egyetem	0.033	0.337	-0.135	1.000

Feladat - 2

Oldja meg a „korreláció - 2” fülön található feladatokat!



Regresszió

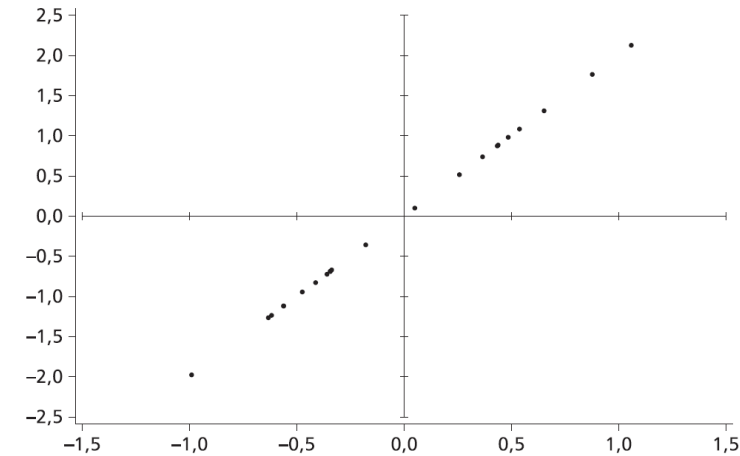


Kapcsolat szorossága grafikusan

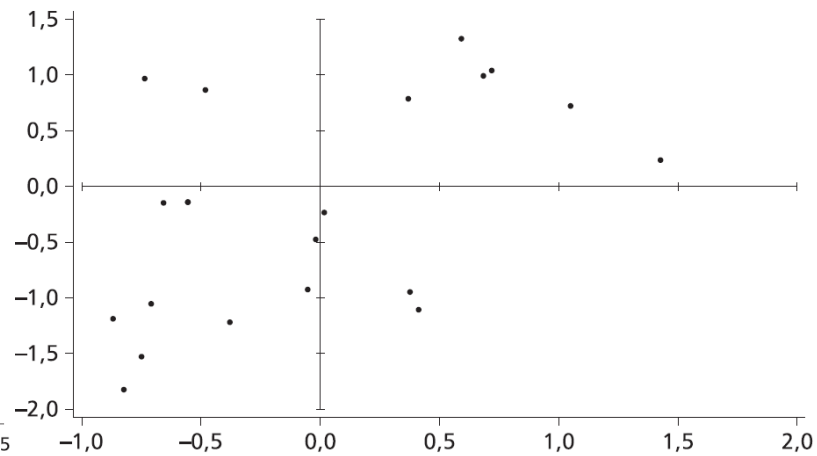
Pontdiagram két változó között

A pontdiagrammon néha nehezen meglátható a kapcsolat.

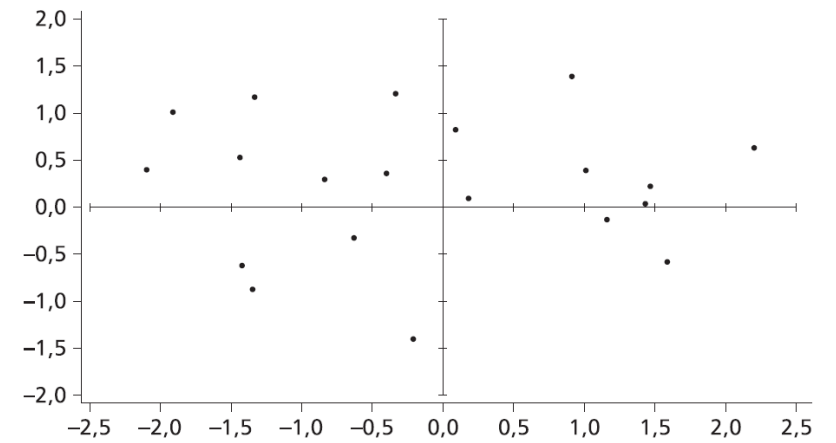
„Mennyire egyszerű egyenest húzni (görbét illeszteni) a pontokra?” És milyen a meredeksége?



3.2. ábra. Két tökéletesen korreláló változó pontdiagramja ($r = 1$)



3.3. ábra. Két pozitívan korreláló változó pontdiagramja ($r = 0,51$)



3.4. ábra. Két egymással nem korreláló (korrelálatlan) változó pontdiagramja ($r = 0$)

Korreláció vs. regresszió

Változók közötti kapcsolat számszerűsítése

Korreláció:

2 változó között

Oksági kapcsolat?

Regresszió:

Komplex összefüggések (több változó és hatásuk)

Lehet mögötte gazdasági modell – okság

Példák: bérregresszió (végzettség, ...), infláció regresszió (kamat, ...), ingatlanár (telek, szobák száma, környék,...)

Egyváltozós regresszió - példa

Ingatlanár és telekméret kapcsolata Windsorban. Ingatlanügynökök vagyunk, ezért érdekel minket. Második okozza az elsőt, ezért az ingatlanár alakulását magyarázzuk telekméret alakulásával. Ingatlanár: függő változó, telekméret: magyarázó (vagy független) változó.

Lépések:

1. Feltesszük, hogy létezik **lineáris kapcsolat** a két változó között. Pl. **egy négyzetlábbal nagyobb telekméret β kanadai dollárral növeli az ingatlan árát**. Minket pontosan β érdekel. Ez eddig egy elméleti modell, amiről úgy gondoljuk, hogy a windsori ingatlanok ára és telekmérete között fennáll. Ez a modell önmagában nem teljes, mert **a telekméreten túl más is hat az ingatlanárakra**. (Például mi?) Azonban a hiba ellenére érdemes a modellel foglalkozni, mert többet tudunk meg az ingatlanárak alakulásáról a telekméret segítségével, mint nélküle.

Egyváltozós regresszió - példa

2. Nem ismerjük az összes windsori ingatlan aktuális árát és telekméretét, csak azokat, amelyeket adott évben eladtak vagy felértékeltek. Az így rendelkezésre álló adatok segítségével próbáljuk megbecsülni az előzőekben felvázolt elméleti modellt. Azaz **csak becsüljük β -t** és reméljük, hogy a becslés valós képet ad, hogy az összes ingatlant tekintve mi az összefüggés az ár és a méret között.
3. Úgy becsüljük a β -t, hogy a telekméret – ingatlanár pontdiagramban szereplő pontokra **ráillesztünk egy olyan egyenest, amely minimalizálja az összes pont (négyzetes) eltérését magától az egyenestől**. Ugyanis a pontok biztosan nem fekszenek egy adott egyenesen, hanem szóródnak körülötte. Minden egyes pontra kiszámoljuk, hogy mennyire tér el az egyenestől (reziduum) és az összeget minimalizáljuk.

Egyváltozós regresszió: az elméleti modell - bevezetés

Y függő változó, X magyarázó változó

Feltevés: lineáris kapcsolat (ha nem az, akkor matematikai transzformációval linearizálható)

Regressziós egyenes:

$$Y = \alpha + \beta X$$

Elméleti modell: feltesszük, hogy X változása hat Y változására, X változása „okozza” Y változását, és a hatás erősségét β méri.

Természetesen nem egy egyenesen fekszenek az adatok (hiszen kihagyunk fontos változókat, amelyek szintén hatnak), minket általános tendenciák érdekelnek.

Egyváltozós regresszió: az elméleti modell - hibatag

Lineáris regresszió: közelítés, azaz hibát követünk el

Kihagyott, meg nem figyelhető változók

Nem lineáris kapcsolat

Regressziós modell hibataggal:

$$Y = \alpha + \beta X + e$$

Hiba: adatpont (ha az összes ingatlanról lenne teljes információnk) és a (valódi) regressziós egyenes közti távolság

Egyváltozós regresszió: a becsült modell

Együtthatók értékeit nem ismerjük, rendelkezésre álló adatokból becsüljük.

Becsült együtthatók: legjobban illeszkedő egyenes együtthatói

Jelölés: $\hat{\alpha}, \hat{\beta}$

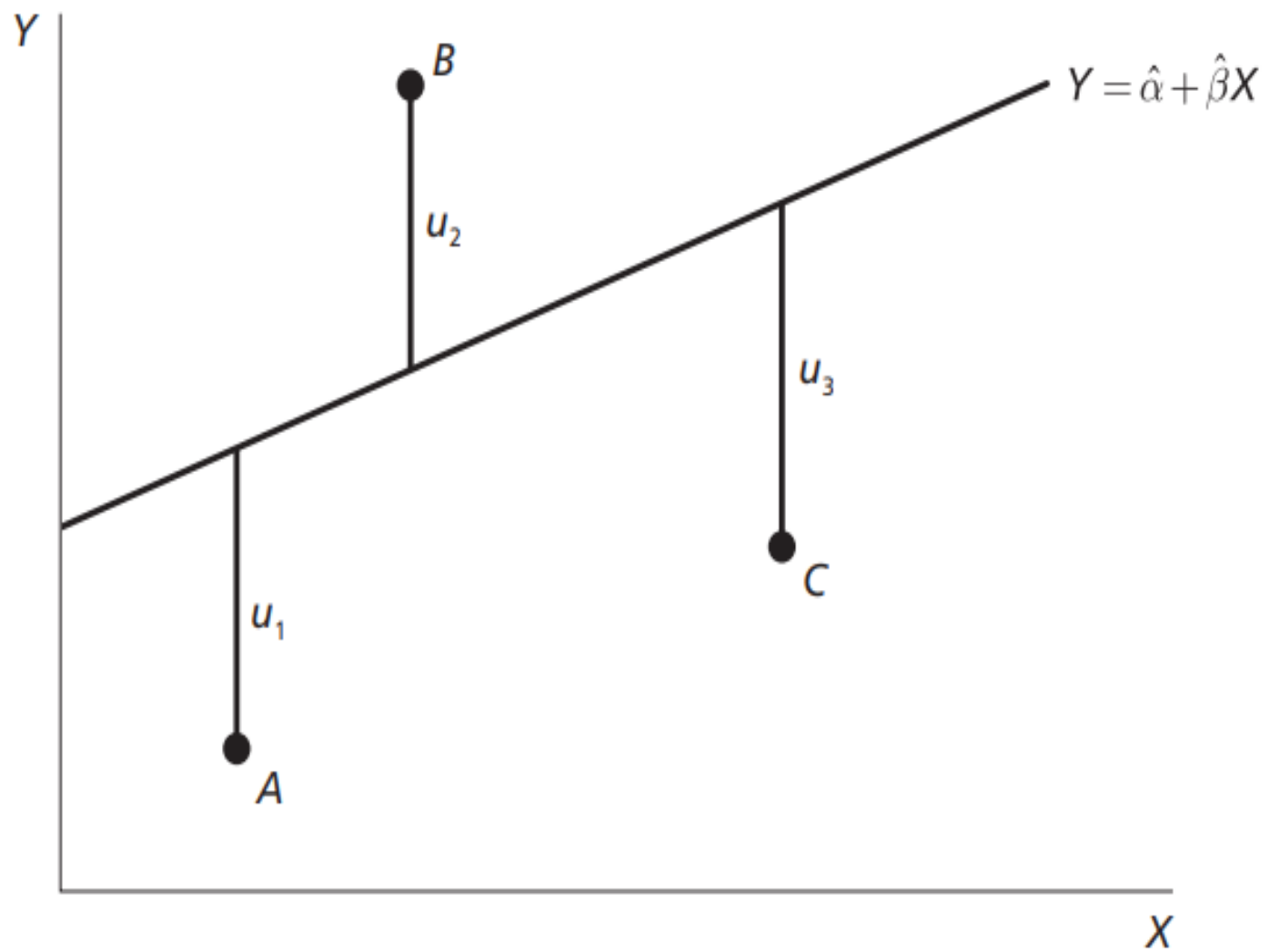
Azaz. $Y = \hat{\alpha} + \hat{\beta}X + u$

Reziduum (maradéktag): $u \neq e$

Meglévő adatpontok nem illeszkednek az egyenesre, egyenes és pont távolsága = reziduum.

Tengelymetszet (konstans) értelmezése – pl. ingatlanárak és telekméret

Egyváltozós regresszió



$\hat{\alpha}$ = tengelymetszet
 $\hat{\beta}$ = meredekség

Az OLS becslés

Legjobban illeszkedő egyenes – maradéktag négyzetösszege (SSR – sum of squared residual) minimális

$$\min SSR = \min \sum_{i=1}^N u_i^2 = \min \sum_{i=1}^N (Y_i - \hat{\alpha} - \hat{\beta}X_i)^2$$

Legkisebb négyzetes becslés = ordinary least squares (OLS)

Ingatlanos példa

Becsült együtthatók:

34136.2 – tengelymetszet;

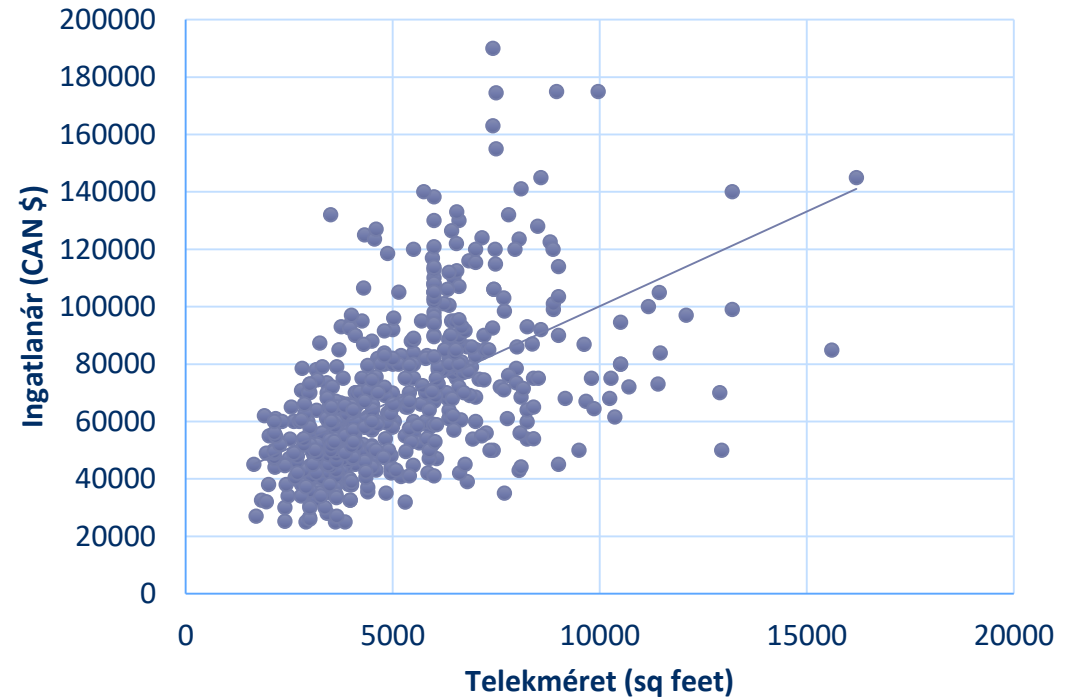
6.6 – telekméret együtthatója

Értelmezés?

Meredekség:

Y átlagos változása X egységnyi növekedése esetén – általános tendenciát ragad meg, nem igaz minden egyes pontra

Marginális hatás: $\frac{dY}{dX} = \beta$



Feladat - 3

Oldja meg a „regresszió” fülön található feladatokat!

Házi feladat

A héten két házifeladatot kell megoldani:

- 1) Egyrészt, az elkezdett Doing Economics fejezet 1.3-as alfejezetét: <https://www.core-econ.org/doing-economics/book/text/01-02.html#part-13-carbon-emissions-and-the-environment>
- 2) Másrészt pedig, a „Bevezetés a Közgazdaságtanba” tárgy óráján elvégzett kísérlet adatainak elemzését. A Moodle-be feltöltött Excel fájlban megtalálhatóak a kérdések. (Az adatbázis a tavalyi csoport adatait tartalmazza.)

A Moodle-be mindkét házi feladatot fel kell tölteni, két külön Excel fájlban.

Köszönöm a figyelmet





Corvinus



Egyváltozós regresszió – illeszkedés, nemlinearitás

Bevezetés az empirikus elemzésbe – 5. hét



Budapesti Corvinus Egyetem
Corvinus University of Budapest

Tartalom



Illeszkedés

8

Nemlinearitás

24

Kitérő 1: Közgazdasági Nobel-díj, 2021



David Card

University of California, Berkeley, USA

*"for his empirical contributions
to labour economics"*



Joshua D. Angrist

Massachusetts Institute of Technology,
Cambridge, USA

*"for their methodological contributions to the analysis
of causal relationships"*



Guido W. Imbens

Stanford University, USA

*„Together, they
have revolutionised
empirical research
in the economic
sciences.”*

- Press release: <https://www.nobelprize.org/uploads/2021/10/press-economicsciencesprize2021.pdf>
- Munkájuk rövid leírása: <https://www.nobelprize.org/uploads/2021/10/popular-economicsciencesprize2021.pdf>

Kitérő 2: 1. negyedéves dolgozat

október 25. - október 31.



1. negyedéves dolgozat előtti kérdések

- 1. negyedéves dolgozat: 2021.10.29., péntek – az órán
- Kérdések – válaszok, majd dolgozat.

I USED TO THINK
CORRELATION IMPLIED
CAUSATION.



THEN I TOOK A
STATISTICS CLASS.
NOW I DON'T.



SOUNDS LIKE THE
CLASS HELPED.

WELL, MAYBE.



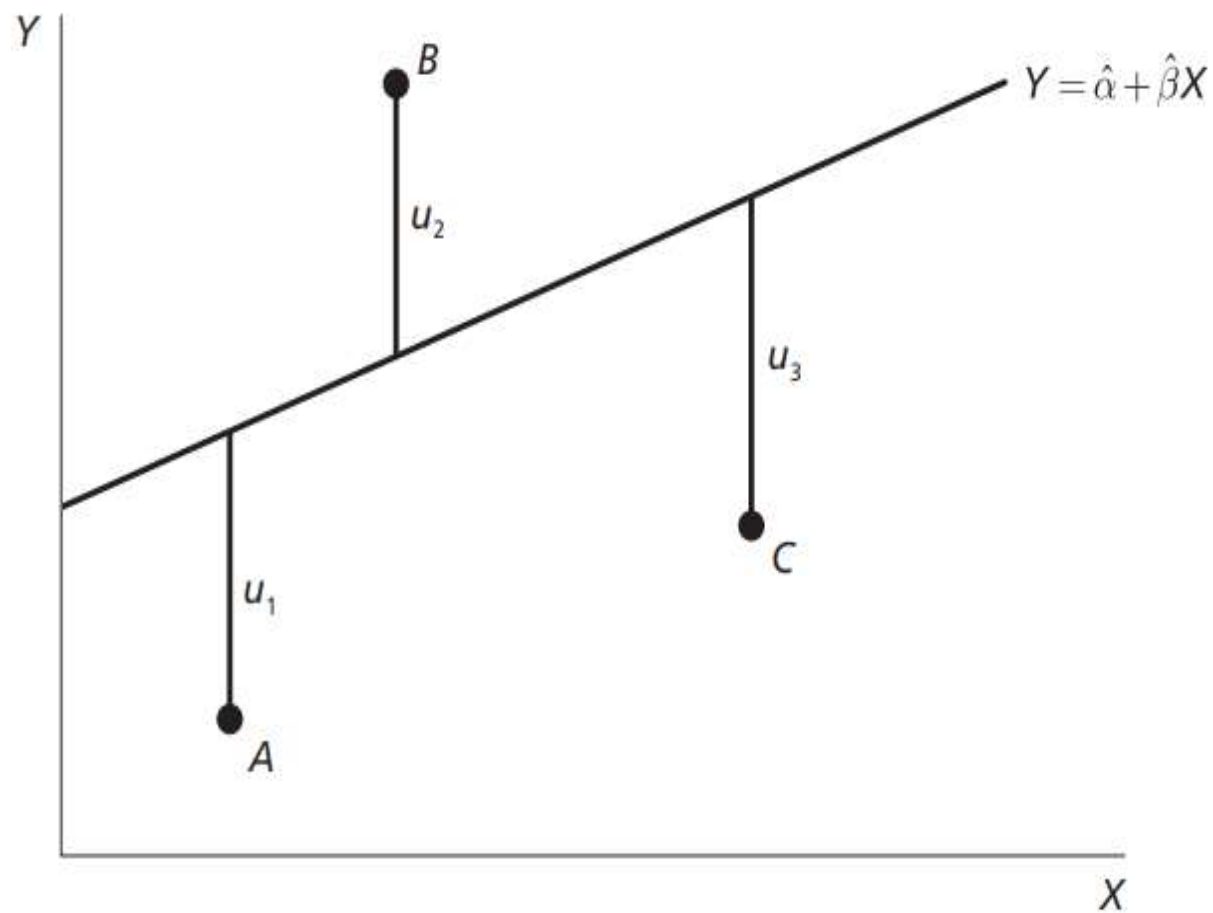


Ismétlés



Egyváltozós regresszió – ismétlés: Lakásár és teleknagyság

- X és Y közötti lineáris kapcsolat:
 - $Y = \alpha + \beta \cdot X$
 - Ez a regressziós egyenes, azaz az X és Y közötti lineáris kapcsolatot leíró összefüggés.
 - Példa: $Y = 35000 + 5 \cdot X$
 - Probléma: kihagyunk fontos változókat → hiba keletkezik.
- A modell: $Y = \alpha + \beta \cdot X + e$
 - A hiba (e): egy adott adatpont és a *valódi* regressziós egyenes közötti távolság.
 - Probléma: α és β értékeit nem ismerjük, azokat csak becsüljük.
- A becsült modell: $Y = \hat{\alpha} + \hat{\beta} \cdot X + u$
- Az OLS becslés: a $\sum u^2$ -t minimalizáljuk.



$\hat{\alpha}$ = tengelymetszet
 $\hat{\beta}$ = meredekség

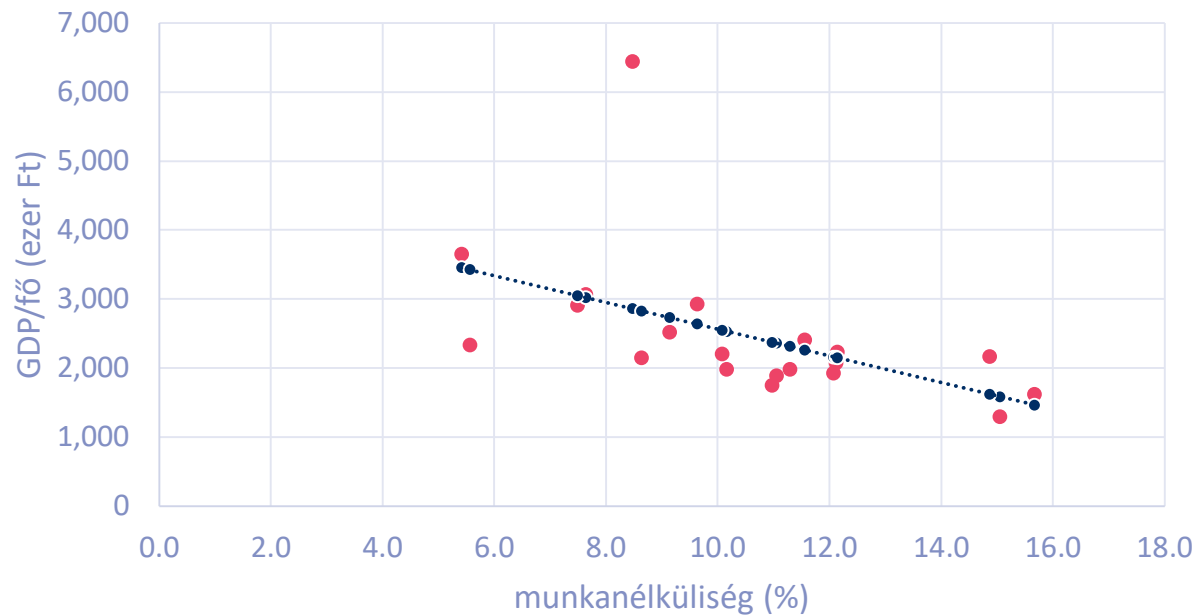
1. példa: munkanélküliség és GDP/fő

Függő változó: GDP/fő (e Ft-ban)

Magyarázóváltozó: munkanélküliségi ráta (%-ban)

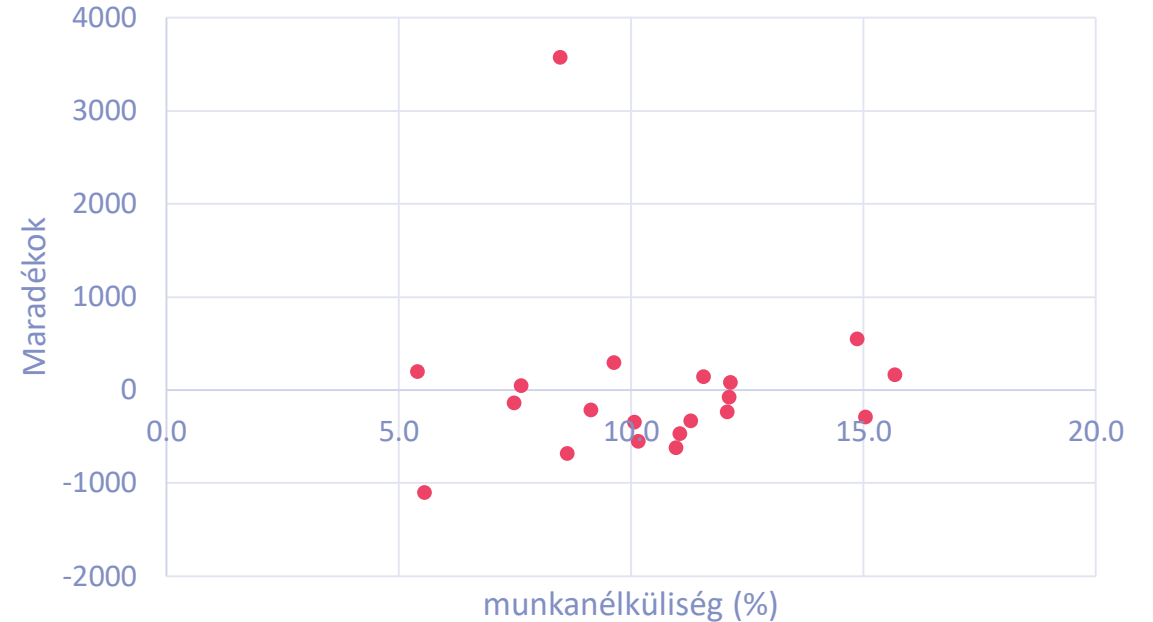
$r^2 = 0.26 \rightarrow$ Értelmezése?

Munkanélküliségi ráta együtthatója: $-194.3 \rightarrow$ Értelmezése?



● Y ● Becsült Y Linear (Becsült Y)

Maradéktagok



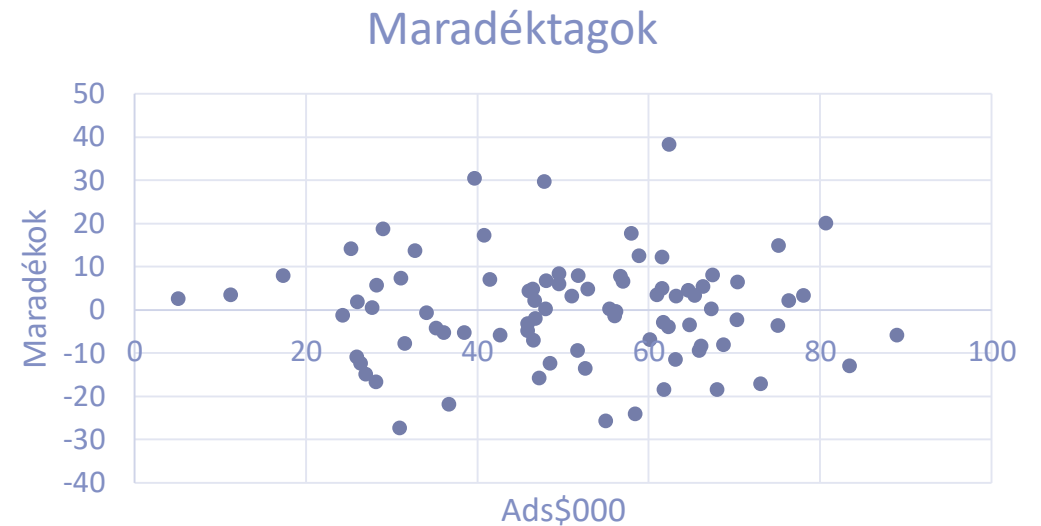
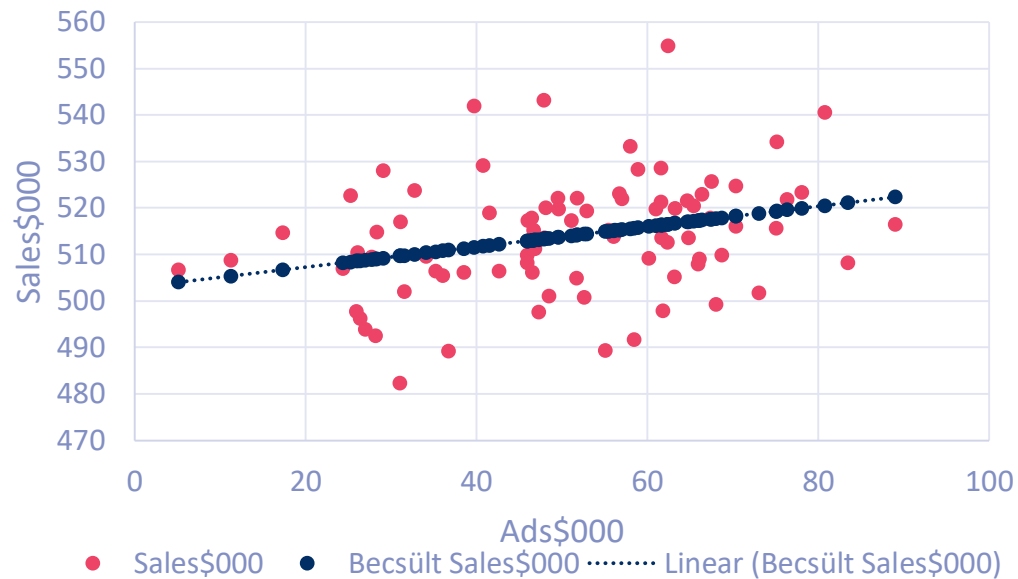
2. példa. hirdetés és értékesítés

Függő változó: értékesítés (e\$-ban)

Magyarázóváltozó: hirdetésre költött összeg (e\$-ban)

$r^2 = 0.09 \rightarrow$ Értelmezése?

Hirdetés együtthatója: 0.22 \rightarrow Értelmezése?



Feladat - 1

Az órai excel fájlban oldja meg a „Forest” fülön található feladatot!



Illeszkedés

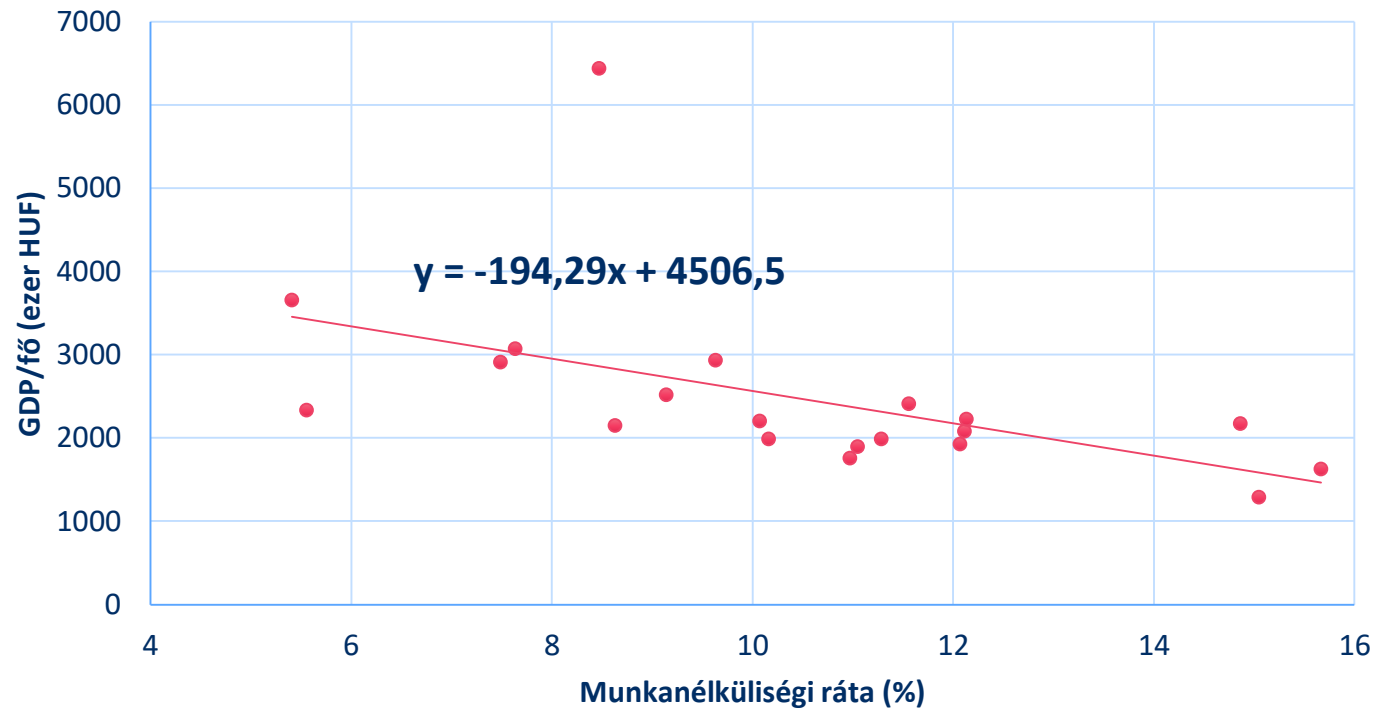


Munkanélküliség és GDP/fő: Mennyire jó a modell?

Függő változó: GDP/fő (e Ft-ban)

Magyarázóváltozó: munkanélküliségi ráta (%-ban)

Kapcsolat az egy főre jutó GDP és a munkanélküliségi ráta között (Magyarország, 2013)



Illeszkedés mérése

OLS: legjobban illeszkedő egyenes megtalálása

Mennyire jó az illeszkedés?

Mérőszám: R^2

Egyváltozós regresszió:

korreláció négyzete = R^2

Illeszkedés – becsült érték

- Regressziós egyenlet: $Y_i = \alpha + \beta X_i + e_i$
- Becsült/illesztett/előrejelzett érték: $\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$
- Kettő összehasonlítása: illeszkedés jósága
- A kulcs: maradéktag $u_i = Y_i - \hat{Y}_i = Y_i - (\hat{\alpha} + \hat{\beta} X_i)$

Azt már láttuk, hogy a regresszió becslésekor a maradéktagok négyzetösszegét (SSR- sum of squared residuals) akarjuk minimalizálni, azaz

$$\min_{\hat{\alpha}, \hat{\beta}} \sum_{i=1}^N (Y_i - \hat{\alpha} - \hat{\beta} X_i)^2$$

Deriválni és nullával egyenlővé tenni.

Illeszkedés – becsült érték

Eredmény

$$\hat{\beta} = \frac{\sum_{i=1}^N (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^N (X_i - \bar{X})^2}$$

$$\hat{\alpha} = \hat{Y}_i - \hat{\beta}X_i$$

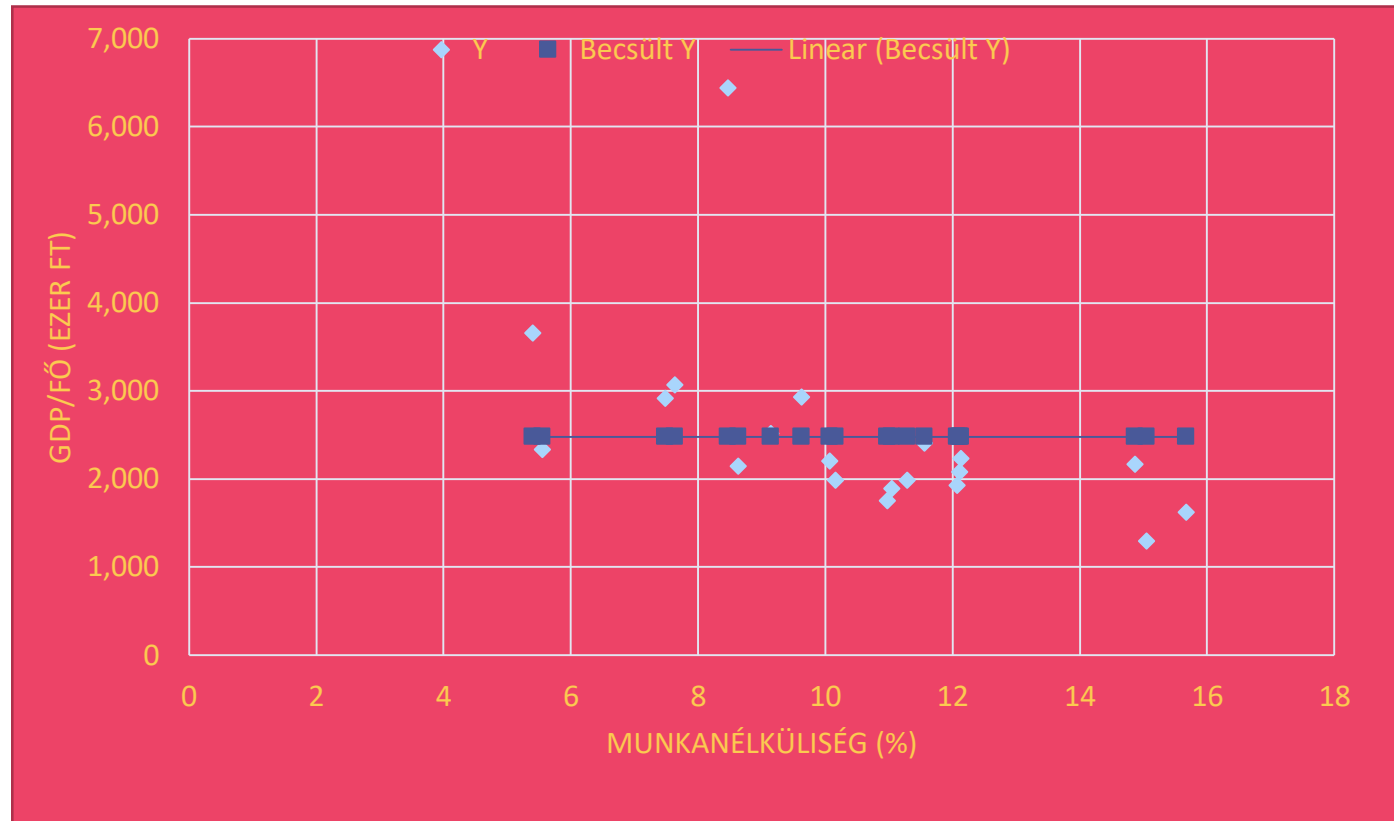
Mire emlékeztet ez?

$$r = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})}{\sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2} \sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}}$$

Jövőre ökonometriából részletek!

Az illeszkedés jósága: R^2

Ha a GDP/fő érdekel minket és semmilyen magyarázó változóval nem rendelkezünk, akkor a legjobb, amit tehetünk az az átlag használata. Azt figyelhetjük meg ekkor, hogy az egyes megfigyelések eltérnek az átlagtól. Ezen eltéréseket akarjuk megérteni a magyarázó változó segítségével.



Az illeszkedés jósága: R^2

Az átlagtól való eltérést a varianciával ragadtuk meg, itt most egy módosított változatot fogunk használni, amit **teljes szórásnégyzet**nek hívunk (TSS, total sum of squares):

$$TSS = \sum_{i=1}^N (Y_i - \bar{Y})^2$$

Tényleges értékek
eltérése az átlagtól.

A varianciához az kellene, hogy $N-1$ -gyel osszuk el. Akár meg is tehetnénk (ha az összes későbbi változóban RSS, SSR is megtennénk). Nem változtatna az eredményen. Inkább mindegyiknél elhagyjuk.

Az SSR-t (maradéktagok négyzetösszege) már ismerjük:

$$SSR = \sum_{i=1}^N u_i^2 = \sum_{i=1}^N (Y_i - \hat{Y}_i)^2$$

Tényleges értékek
eltérése a becsült
értékektől.

Az illeszkedés jósága: R^2

A TSS és az SSR között a kapcsolatot a regressziós négyzetösszeg (RSS, regression sum of squares) teremti meg.

Az RSS nem más, mint

$$RSS = \sum_{i=1}^N (\hat{Y}_i - \bar{Y})^2$$

Becsült értékek
eltérése az átlagtól.

Azt mutatja, hogy a regresszió által becsült értékek mennyire térnek el az átlagtól és ezen eltérések négyzetösszegét adja meg az RSS.

Alapötlet: Átlaghoz képest a megfigyelések szóródnak. Regresszióhoz képest is szóródnak a megfigyelések (maradékok), de jóval kevésbé, mint az átlaghoz képest. Azaz a regresszió „megfogja” a megfigyelések szóródásának egy részét. Úgy számszerűsítjük, hogy megnézzük a megfigyelések átlaghoz viszonyított szóródását és a megbecsült értékek átlaghoz mért szóródását. Ha a kettő közel van egymáshoz, akkor a megfigyelések szóródását a regresszió jól magyarázza.

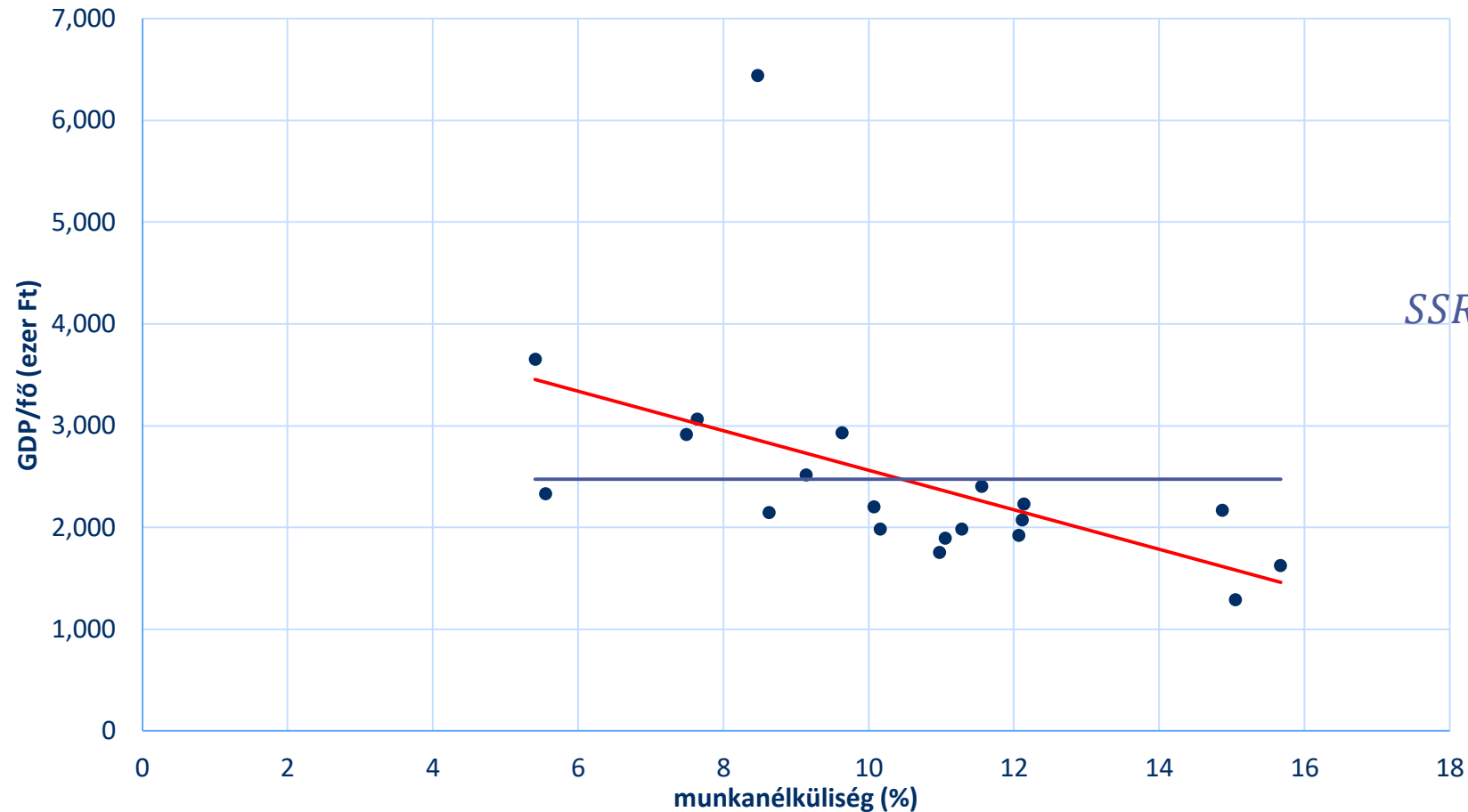
Az illeszkedés jósága: R^2

Belátható, hogy $TSS=RSS+SSR$

$$\begin{aligned}\sum_{i=1}^N (Y_i - \bar{Y})^2 &= \sum_{i=1}^N (Y_i - \hat{Y}_i + \hat{Y}_i - \bar{Y})^2 = \\ \sum_{i=1}^N ((Y_i - \hat{Y}_i) + (\hat{Y}_i - \bar{Y}))^2 &= \sum_{i=1}^N (u_i + (\hat{Y}_i - \bar{Y}))^2 = \\ \sum_{i=1}^N (u_i^2 + (\hat{Y}_i - \bar{Y})^2 + 2u_i(\hat{Y}_i - \bar{Y})) &= \\ \sum_{i=1}^N u_i^2 + \sum_{i=1}^N (\hat{Y}_i - \bar{Y})^2 + 2\sum_{i=1}^N u_i(\hat{Y}_i - \bar{Y}) &= \\ \sum_{i=1}^N u_i^2 + \sum_{i=1}^N (\hat{Y}_i - \bar{Y})^2 &= SSR + RSS\end{aligned}$$

Utolsó előtti sorban utolsó tag azért esik ki, mert u_i és $\hat{Y}_i - \bar{Y}$ nem mozognak együtt (ha igen, akkor rossz a regresszió) és a szorzatuk összege az összes megfigyelést figyelembe véve nulla.

Az illeszkedés jósága: R^2



$$TSS = \sum_{i=1}^N (Y_i - \bar{Y})^2$$

$$SSR = \sum_{i=1}^N u_i^2 = \sum_{i=1}^N (Y_i - \hat{Y}_i)^2$$

$$RSS = \sum_{i=1}^N (\hat{Y}_i - \bar{Y})^2$$

TSS: a megfigyelések szóródása az átlaghoz viszonyítva

RSS: a regresszió segítségével becsült pontok szóródása az átlaghoz képest

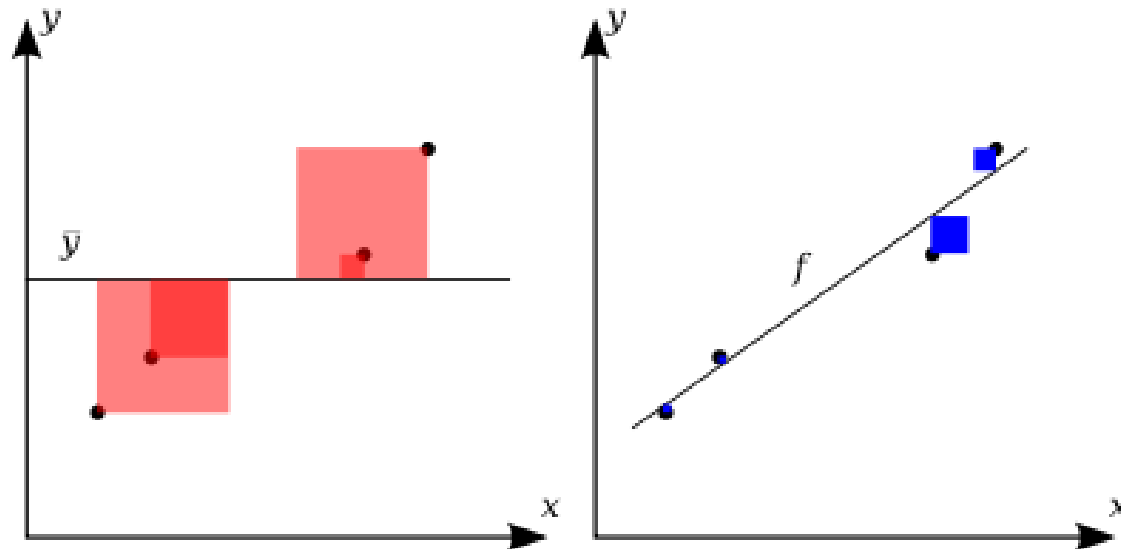
SSR: a megfigyelések szóródása a regressziós egyeneshez képest

R² másként

Bal oldalon 4 pont és az átlaguk, valamint az átlagtól való négyzetes eltérés látható négyzetekkel ábrázolva. - TSS

Jobb oldalon a 4 pontra illesztett regressziós egyenes látható és a megfigyelések (pontok) egyenestől való négyzetes eltérése. – SSR

Az R² azt mutatja, hogy a nagy piros négyzetek összege hogyan megy össze a regresszió segítségével és lesz kis kék négyzetek összege.



Az illeszkedés jósága: R^2

R^2 azt mutatja, hogy a regresszió a teljes variancia hány százalékát magyarázza meg. Számszerűleg:

$$R^2 = 1 - \frac{SSR}{TSS} = \frac{RSS}{TSS}$$

Látszik, hogy ha RSS-t és TSS-t is leosztanánk $N-1$ -gyel, akkor ugyanezt az eredményt kapnánk. Azaz a fenti kifejezés tényleg azt mutatja, hogy a teljes szórásnégyzet hány százalékát magyarázza a „regressziós szórásnégyzet” (ilyen nincs, csak RSS).

R² összefoglaló

Teljes szórásnégyzet:

$$TSS = \sum_{i=1}^N (Y_i - \bar{Y})^2$$

Variancia = TSS/(N-1)

Regressziós négyzetösszeg:

$$RSS = \sum_{i=1}^N (\hat{Y}_i - \bar{Y})^2$$

Maradéktag négyzetösszeg:

$$SSR = \sum_{i=1}^N (Y_i - \hat{Y}_i)^2$$

Illetve

$$TSS = RSS + SSR$$

és

$$R^2 = 1 - \frac{SSR}{TSS} = \frac{RSS}{TSS}$$

R² értelmezése

Y varianciájából hány % magyarázható X segítségével

$$0 \leq R^2 \leq 1$$

Ha $R^2=1$, akkor tökéletes az illeszkedés.

Nullához közeli R^2 rossz illeszkedésre utalhat, de alacsony R^2 nem jelenti azt, hogy egy regresszió rossz.

R² értelmezése

Mindent összevetve, az R^2 igen sok összefüggésben megjelenő, szinte központi fontosságú diagnosztikai mutató, még akkor is, ha az utóbbi években szerepe átértékelődött. Természetesen nem szabad fetisizálni, de kellő óvatossággal használni lehet és kell, hiszen nagymértékben segíti a modellépítés munkáját. Nagyon kicsi értékei veszélyeket hordoznak, ezért bár a nagy R^2 önmagában nem jelenti azt, hogy a modell jó, a kicsi nagy valószínűséggel azt jelzi, hogy a modell rossz. Lehet, hogy a kicsi néha szép, de többnyire nem jó.

Forrás: Hunyadi László (2000): A determinációs együtthatóról. Statisztikai Szemle 78(9): 753-765.

Feladat - 2

Az órai excel fájlban oldja meg a „R2” fülön található feladatot!



Nemlinearitás



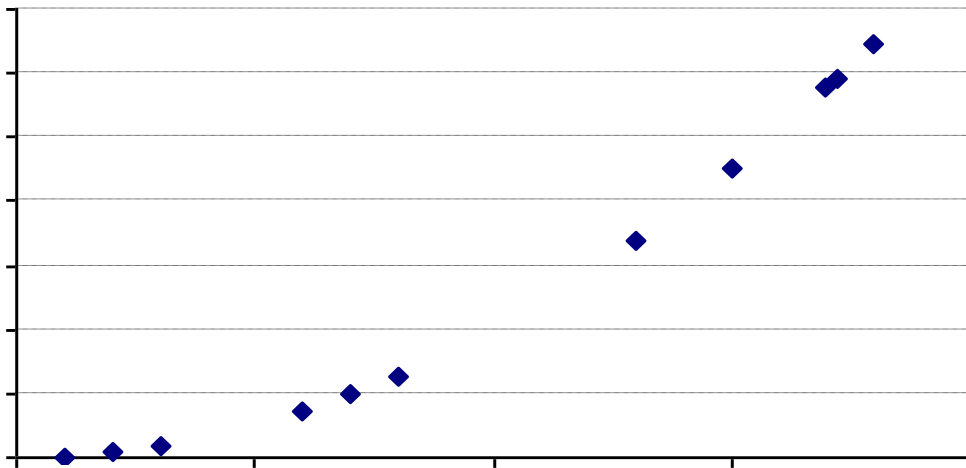
Nemlineáris kapcsolat

Feltevésünk: X és Y közötti *lineáris* kapcsolat van, vagyis $Y = \alpha + \beta \cdot X$

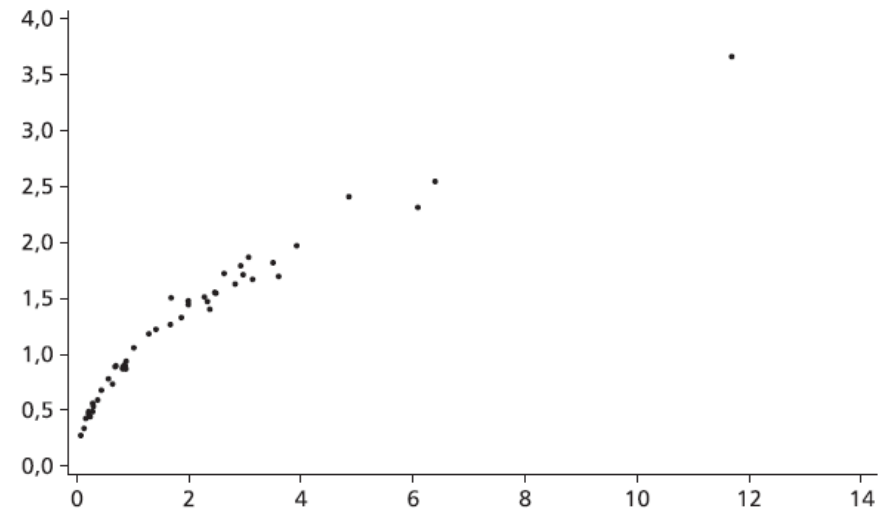
A valóságban a kapcsolat X és Y között gyakran nem lineáris

Gyakori példák:

Négyzetes:



Logaritmikus:



Megoldás nemlinearitás esetén

Olyan matematikai transzformációt alkalmazni, amely linearizálja a kapcsolatot.

Függvények: kvadratikus, logaritmus stb...

Nemlinearitás gyakori lehetséges közgazdasági oka: csökkenő határhaszon.

Példa: bér hogyan függ a munkatapasztalattól (években megadva)

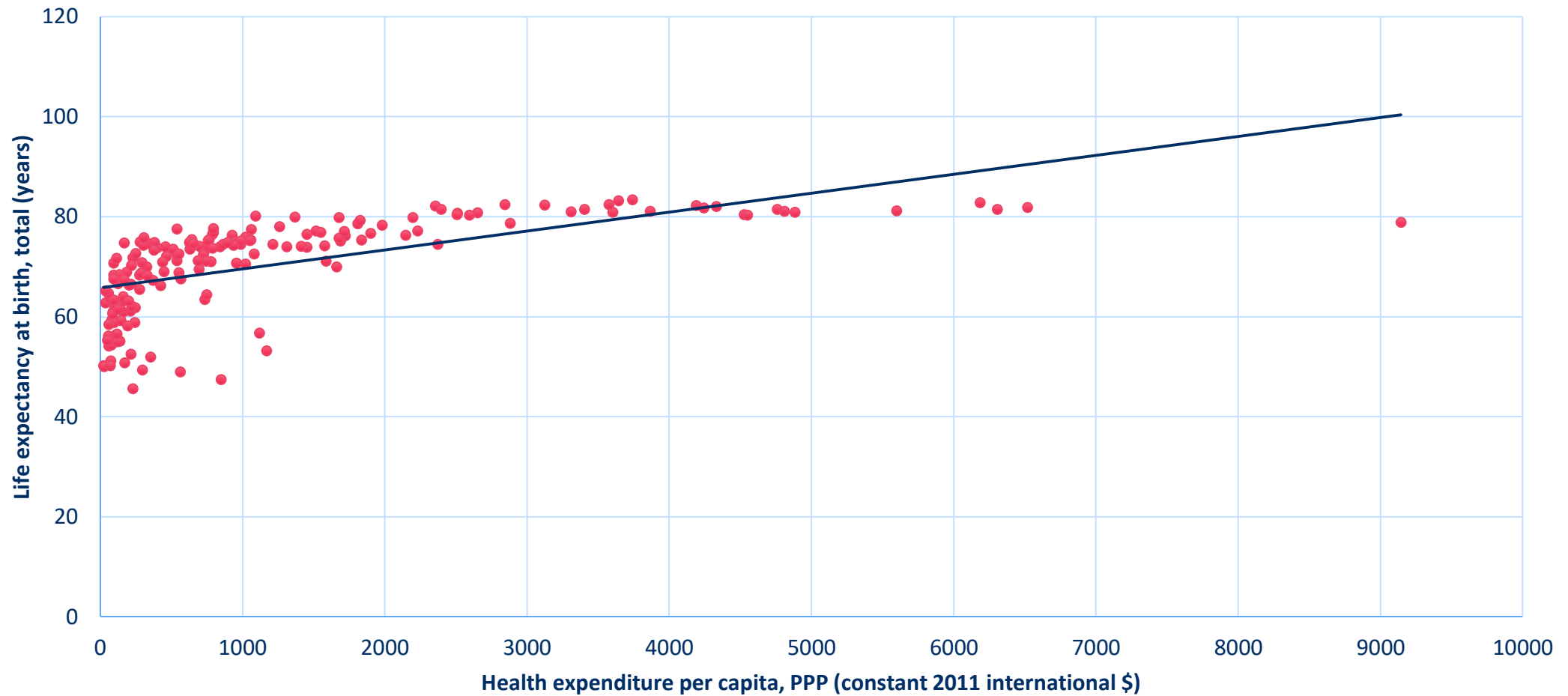
$$bér = 5,25 + 0,48tap - 0,008tap^2$$

Feladat - 3

Az órai excel fájlban oldja meg a „Kvadratikus” fülön található feladatot!

Példa nemlinearitás

2013-as adatok várható élettartamra és egészségügyi kiadások/fő (forrás: World Bank Databank)



Példa nemlinearitás

Látszik, hogy az összefüggés nem lineáris. Ha ráillesztünk simán egy lineáris regressziót, akkor ezt kapjuk:

Sima regresszió

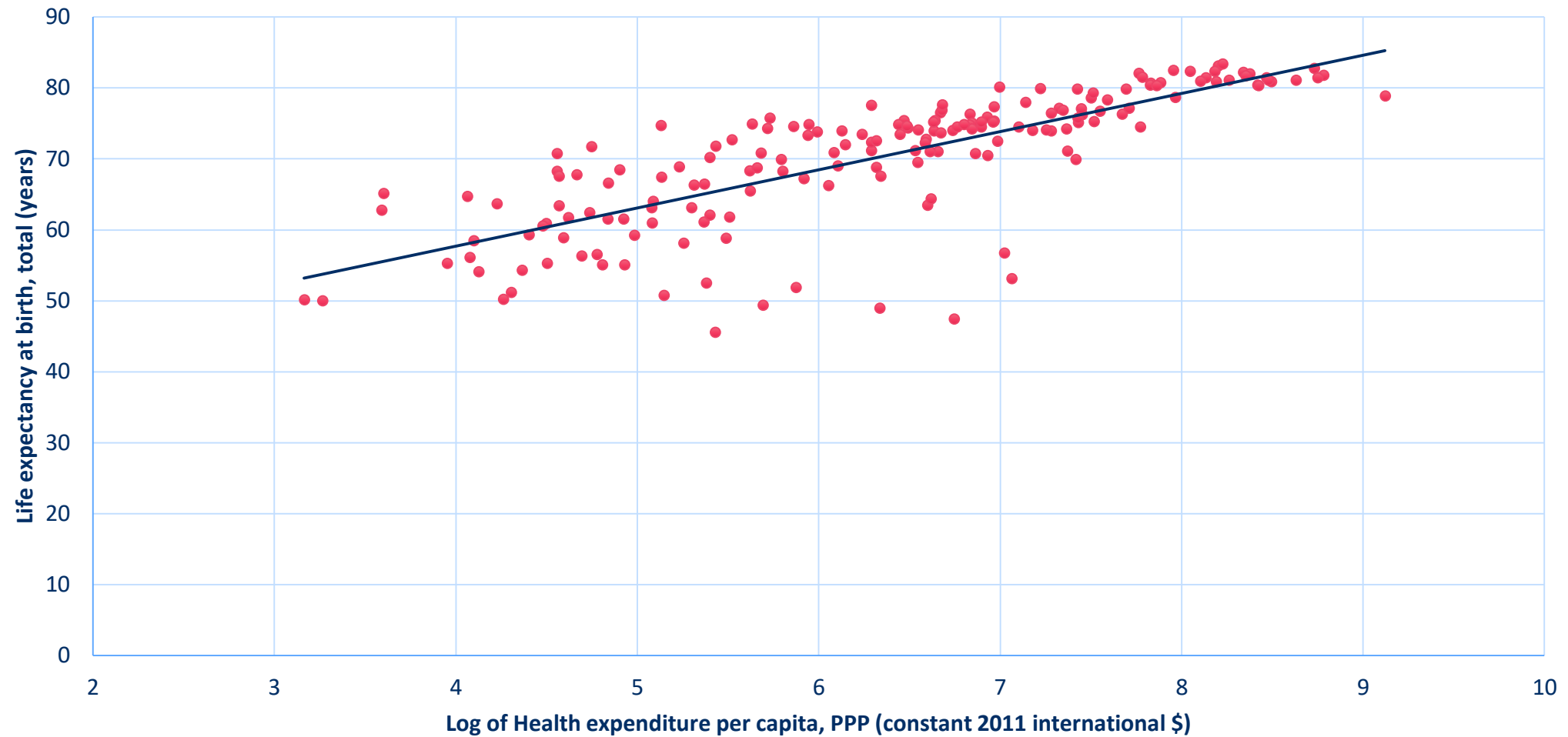
ÖSSZESÍTŐ TÁBLA

<i>Regressziós statisztika</i>	
r értéke	0.624364
r-négyzet	0.389831
Korrigált r-négyzet	0.386403
Standard hiba	7.149877
Megfigyelések	180

	<i>Koefficiensek</i>	<i>Standard hiba</i>	<i>t érték</i>	<i>p-érték</i>	<i>Alsó 95%</i>	<i>Felső 95%</i>
Tengelymetszet	65.80181	0.683256	96.30621	1.7E-155	64.45348	67.15013
Health expenditure per capita, PPP (constant 2011 international \$)	0.003777	0.000354	10.66407	7.62E-21	0.003078	0.004476

Példa nemlinearitás

Vegyük az egészségügyi kiadások logaritmusát!



Példa nemlinearitás

Új regresszió most a kiadások logaritmusával

Log(health) regresszió

ÖSSZESÍTŐ TÁBLA

<i>Regressziós statisztika</i>	
r értéke	0.781555
r-négyzet	0.610828
Korrigált r-négyzet	0.608642
Standard hiba	5.710105
Megfigyelések	180

	<i>Koefficiensek</i>	<i>Standard hiba</i>	<i>t érték</i>	<i>p-érték</i>	<i>Alsó 95%</i>	<i>Felső 95%</i>
Tengelymetszet	36.21021	2.087055	17.34991	4.09E-40	32.09166	40.32876
log (health exp)	5.38104	0.321934	16.71471	2.54E-38	4.745741	6.016339

Javult a regresszió illeszkedése?

Logaritmikus forma

Lineáris összefüggést eredményezhet.

Könnyű értelmezhetőség – rugalmasság:

$$\ln Y = \alpha + \beta \ln X$$

$$\beta = \frac{d \ln Y}{d \ln x}$$

X egy %-os változásakor Y általában β %-kal változik

X adott értékének nincs szerepe (nemlinearitás nem gond).

Mértékegységnek nincs szerepe.

Klasszikus példa: árrugalmasság – ha egy %-kal nő az ár, hány %-kal változik a kereslet.

Példa: jövedelem vs. év

Logaritmiikus forma

Ha x értéke r %-kal nő, akkor x -ből $x(1+r)$ lesz.

Ez a változás logaritmusban kifejezve

$$\log(x(1+r)) - \log(x) = \log(x) + \log(1+r) - \log(x) = \log(1+r).$$

$\log(1+r) \approx r$, ha r kicsi.

r	$\log(1+r)$
-50%	-0.69
-40%	-0.51
-30%	-0.36
-20%	-0.22
-10%	-0.11
-5%	-0.05
-2%	-0.02
0%	0.00
2%	0.02
5%	0.05
10%	0.10
20%	0.18
30%	0.26
40%	0.34
50%	0.41

Szint, log, regresszió

Szint-szint regresszió (level-level):

$$Y_i = \alpha + \beta X_i + e_i$$

„Ha X egy egységgel változik, akkor azt várjuk, hogy Y β egységgel változzon.”

Log-szint regresszió (log-level):

$$\ln Y_i = \alpha + \beta X_i + e_i$$

„Ha x egy egységgel változik, akkor azt várjuk, hogy Y $100 \cdot \beta$ százalékkal változzon.”

Szint-log regresszió (level-log):

$$Y_i = \alpha + \beta \ln X_i + e_i$$

„Ha X egy százalékkal változik, akkor azt várjuk, hogy Y $(\beta/100)$ egységgel változzon.”

Log-log regresszió (log-log):

$$\ln Y_i = \alpha + \beta \ln X_i$$

„Ha X egy százalékkal változik, akkor azt várjuk, hogy Y β százalékkal változzon.”

TABLE 2.3

Summary of Functional Forms Involving Logarithms

Model	Dependent Variable	Independent Variable	Interpretation of β_1
Level-level	y	x	$\Delta y = \beta_1 \Delta x$
Level-log	y	$\log(x)$	$\Delta y = (\beta_1/100)\% \Delta x$
Log-level	$\log(y)$	x	$\% \Delta y = (100\beta_1) \Delta x$
Log-log	$\log(y)$	$\log(x)$	$\% \Delta y = \beta_1 \% \Delta x$

Feladat - 4

Az órai excel fájlban oldja meg a „Nemlinearitás” fülön található feladatot!

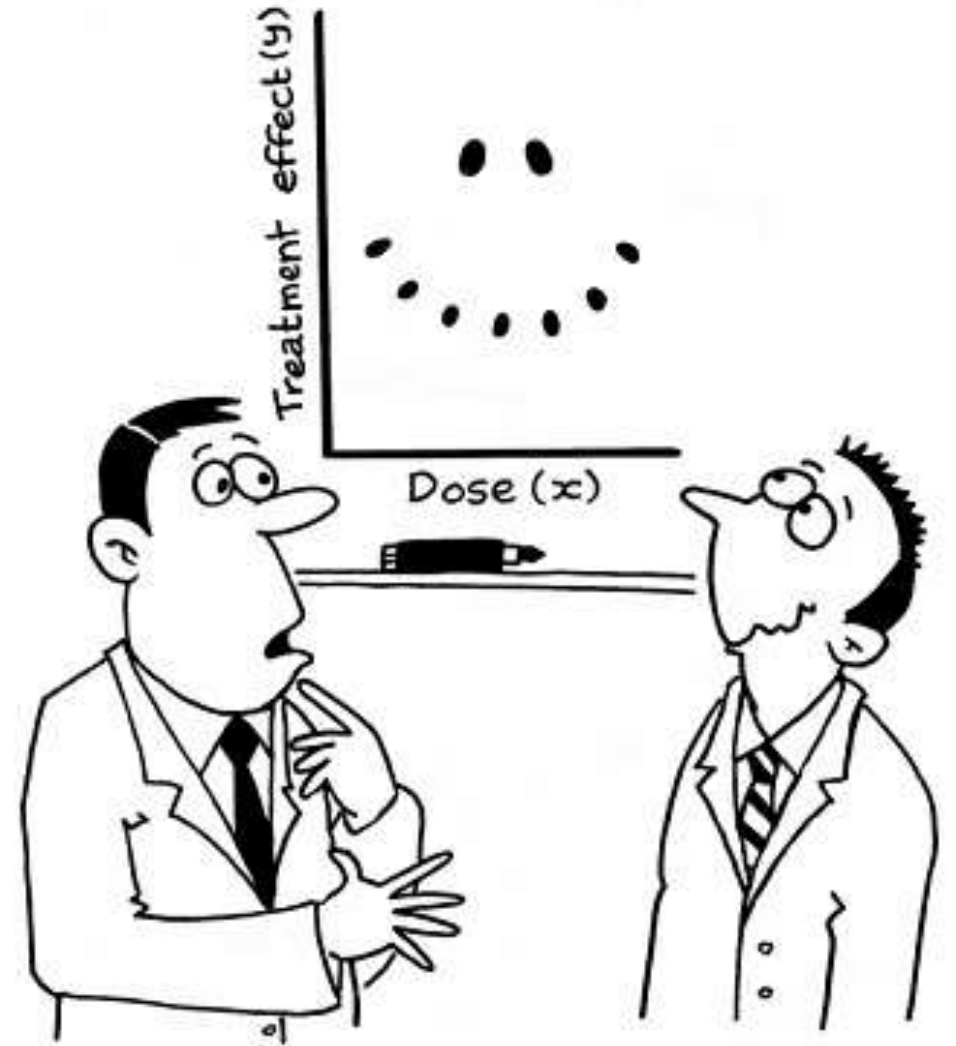
Összefoglalás

Zárózene: [Correlation song](#)



Are points near a line, or far?
What's the correlation, r ?
If the fit supports a line,
Its slope and r would share the sign.
Twinkle, twinkle, you're a star:
Knowing stats will take you far!

Forrás: <https://online.stat.psu.edu/stat100/lesson/5/5.7>



"It's a non-linear pattern with outliers.....but for some reason I'm very happy with the data."

Házi feladat

A Moodle platformra feltöltve Excel fájlban a tankönyvből már jól ismert hirdetés-eladás adatbázist találja. Kövesse az ott megadott instrukciókat!

Köszönöm a figyelmet





Corvinus



Konfidencia intervallum

Bevezetés az empirikus elemzésbe – 6. hét



Budapesti Corvinus Egyetem

Corvinus University of Budapest

Tartalom



Bizonytalanság

Konfidencia intervallum

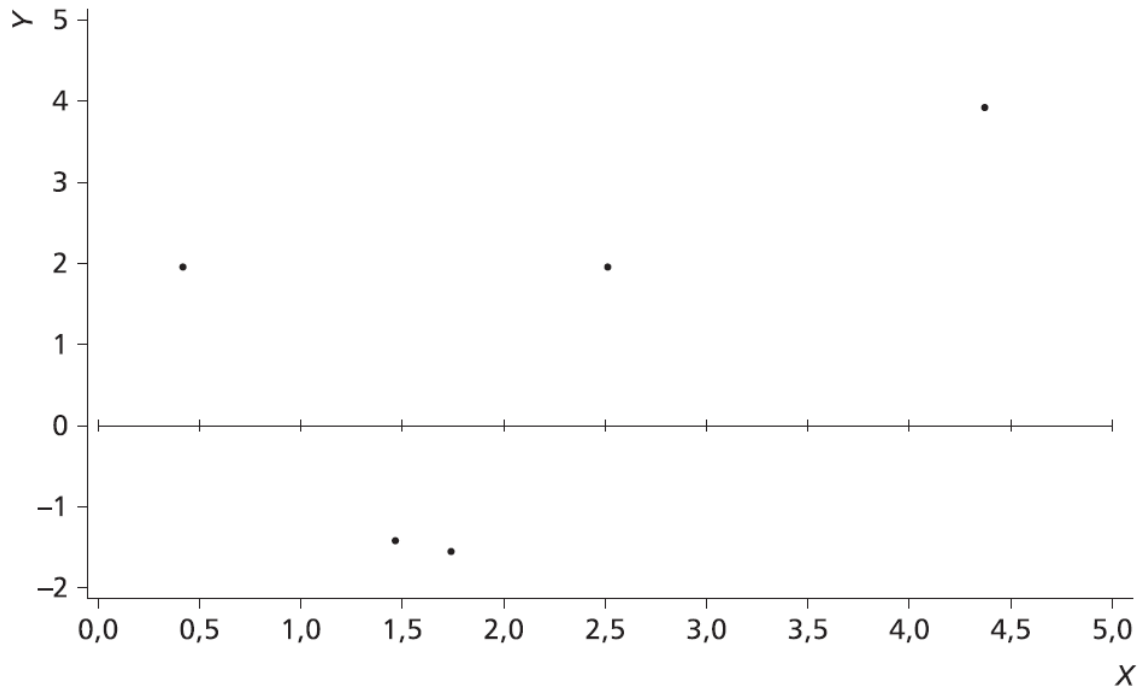
~~Hipotézis vizsgálat~~



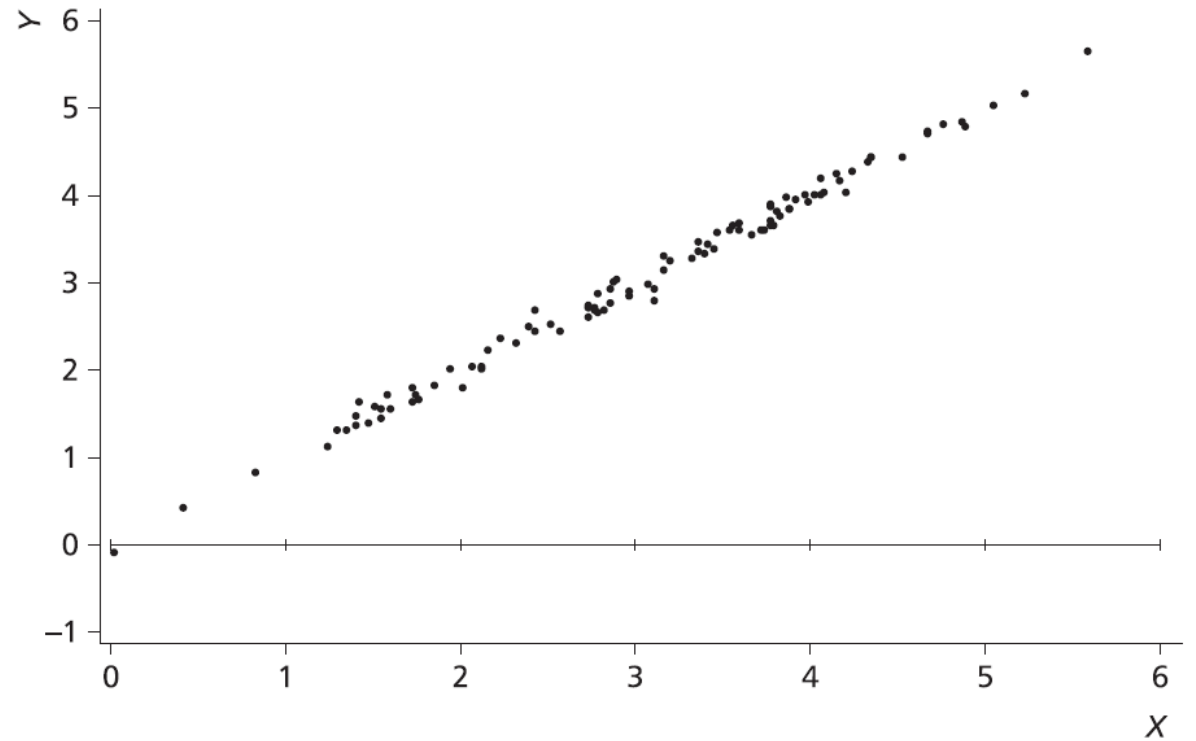
Bizonytalanság



Bizonytalanság (1): Több adatpont – pontosabb becslés: egyre több pont, egyre pontosabb képet ad, azaz bizonytalanság csökken.

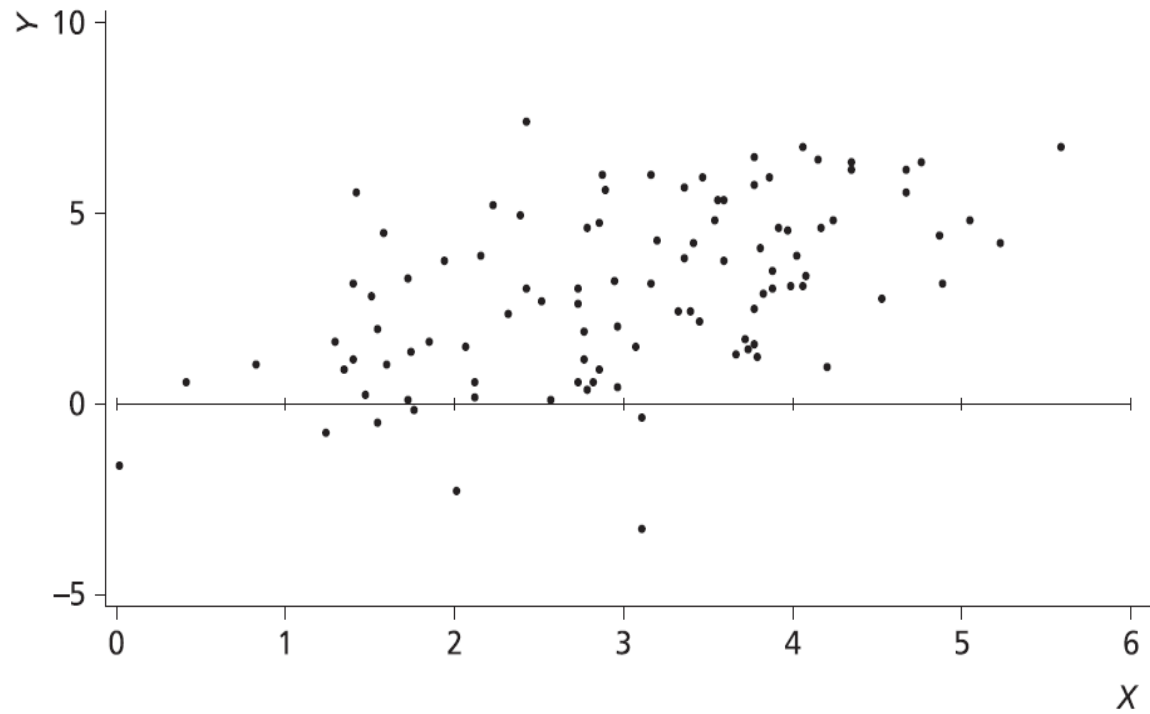


5.1. ábra. Igen kis minta

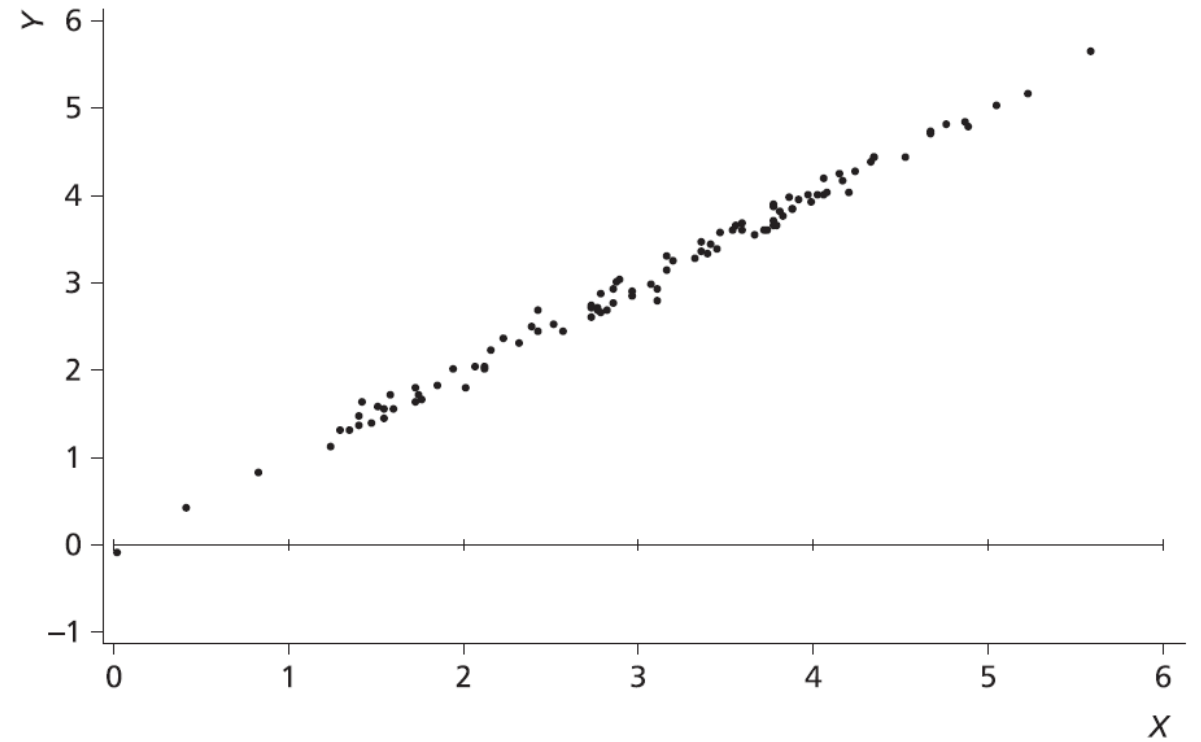


5.3. ábra. Nagy minta, a hibák varianciája kicsi

Bizonytalanság (2): Kisebb hibatagok – pontosabb becslés.

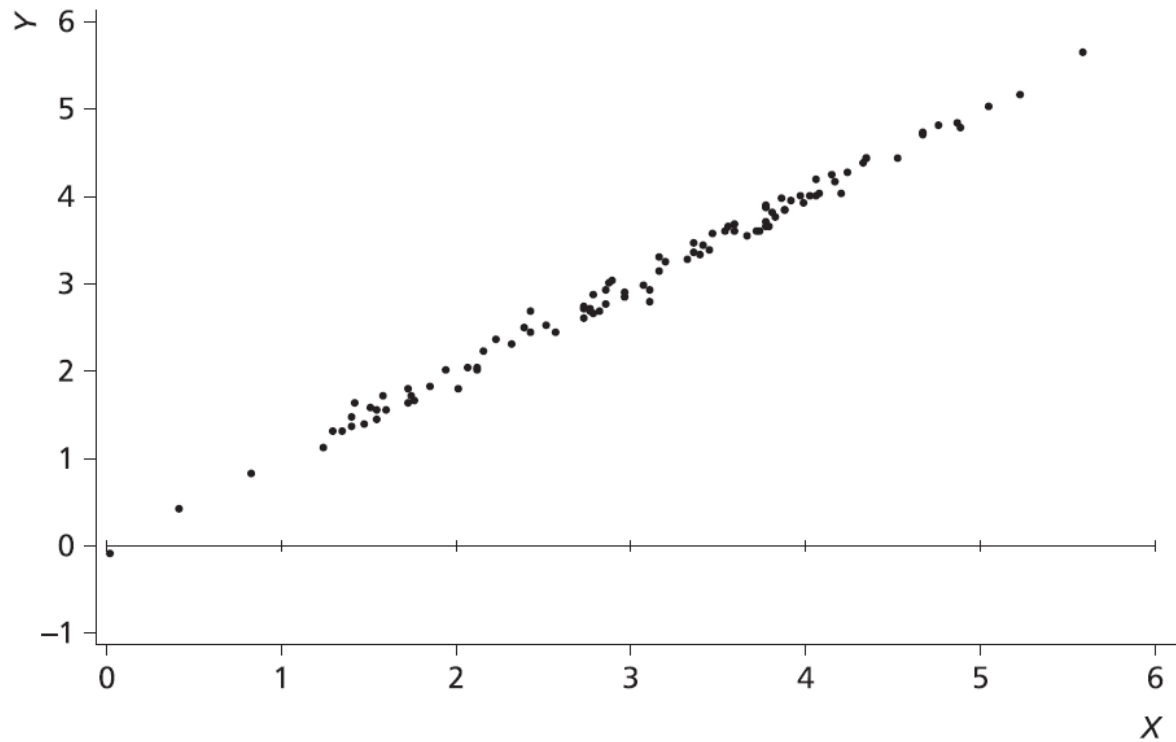


5.2. ábra. Nagy minta, a hibák varianciája nagy

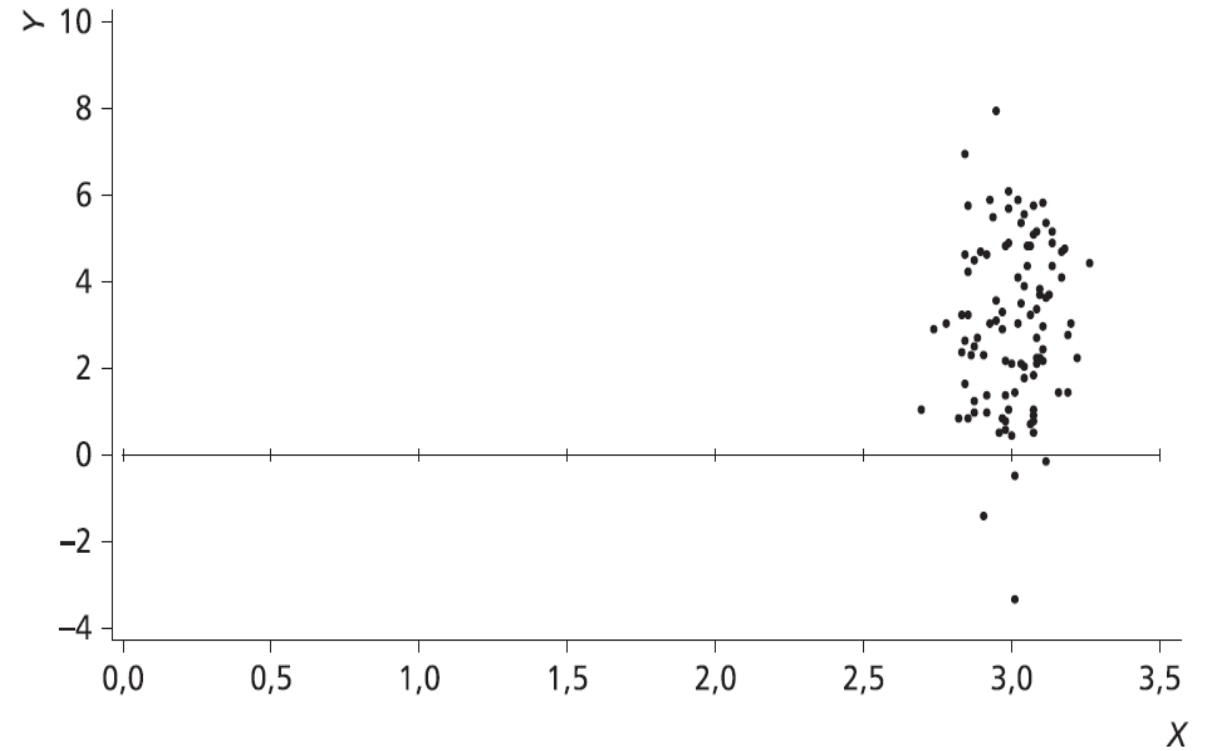


5.3. ábra. Nagy minta, a hibák varianciája kicsi

Bizonytalanság (3): X nagyobb szóródása – pontosabb becslés: az információ a változékonyságban van, X változása milyen hatást generál Y -ban.



5.3. ábra. Nagy minta, a hibák varianciája kicsi



5.4. ábra. Az X értékeinek terjedelme kicsi

Bizonytalanság

(1) Több adatpont – pontosabb becslés: egyre több pont, egyre pontosabb képet ad, azaz bizonytalanság csökken.

(2) Kisebb hibatagok – pontosabb becslés.

(3) X nagyobb szóródása – pontosabb becslés: az információ a változékonyságban van, X változása milyen hatást generál Y -ban.

Példa: végzettség hatásának becslése jövedelemre

Feladat - 1

Az órai Excel fájlban oldja meg a „Beta” fülön található feladatokat!



Konfidencija intervallum



Hány dekagramm az átlag?



11,7 dkg

11,2 dkg



Bizonytalanság

Regressziós együtthatók valódi értéke (azaz α és β) nem ismert: $Y = \alpha + \beta \cdot X + e$

Minta alapján becsülünk: $Y = \hat{\alpha} + \hat{\beta} \cdot X + u$

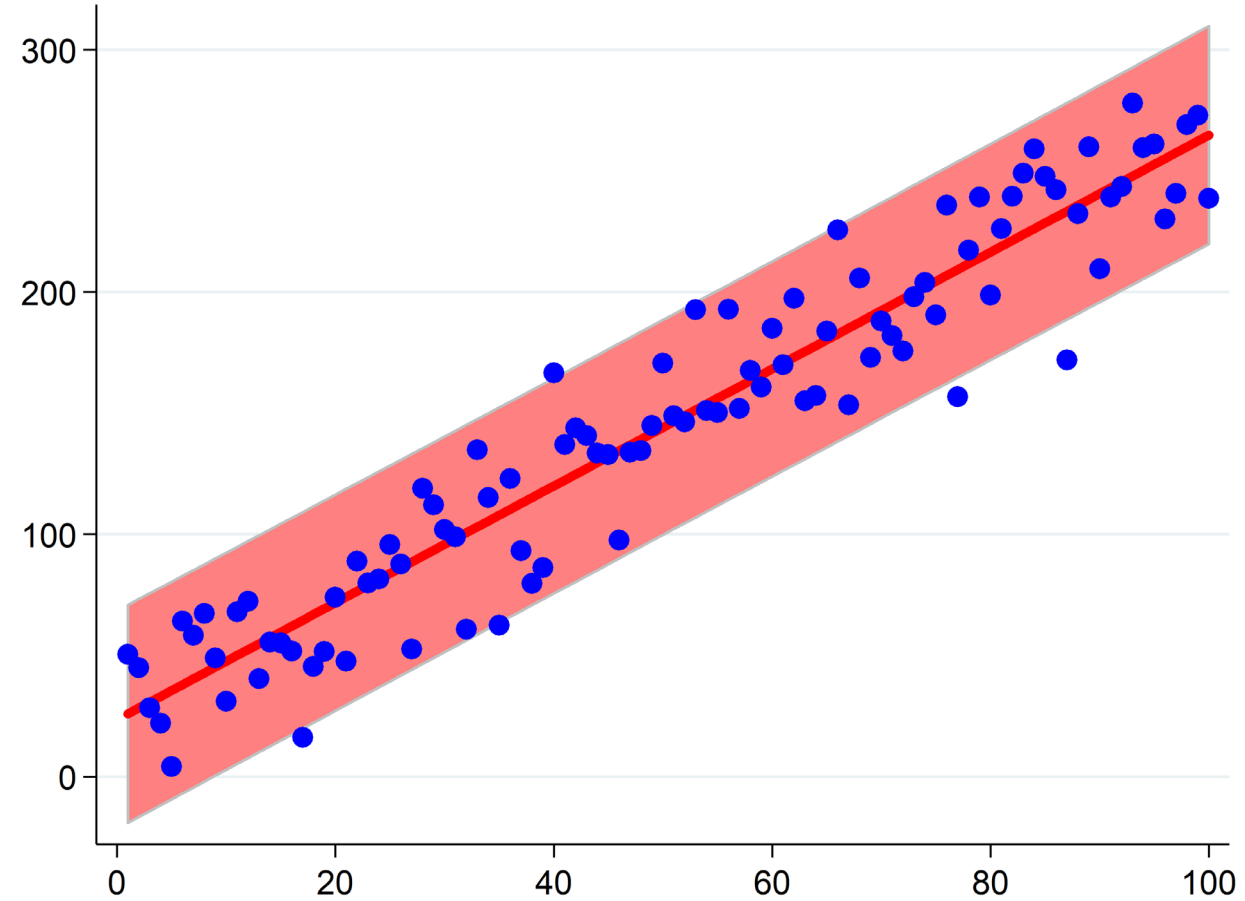
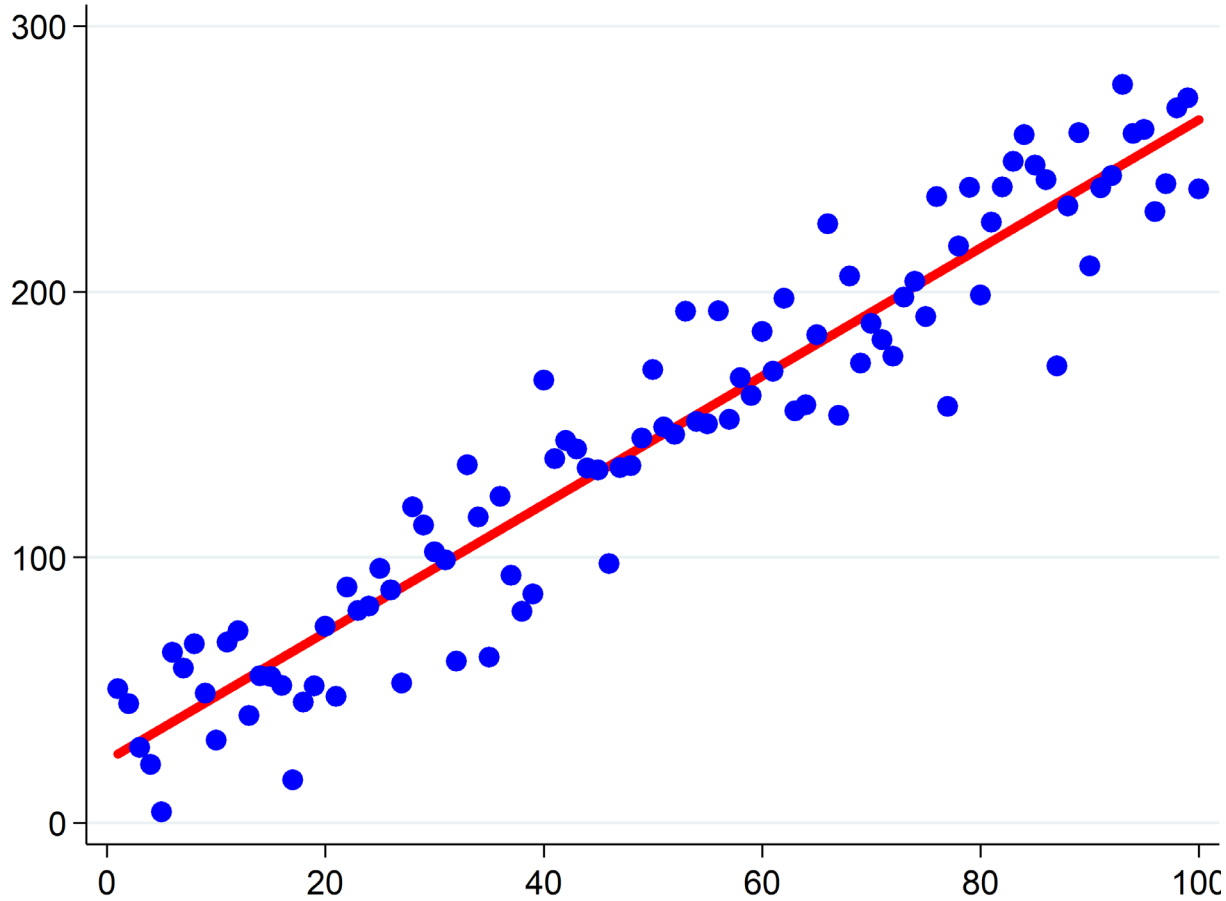
Becsült érték nem pontosan azonos a valódi értékkel. Mennyire *megbízható* a becslés?

A becslés után *pontbecslésünk* lesz. Ez a β -ra adható legjobb közelítés, de a bizonytalanságot nem tükrözi. (u és GDP/fő: -194.3; hirdetés és értékesítés: 0.22) \rightarrow pontbecslés = β értékére adható legjobb közelítés.

Két módszer arra, hogy megnézzük, hogy a becsült érték mennyire pontos, illetve mennyire különbözik nullától:

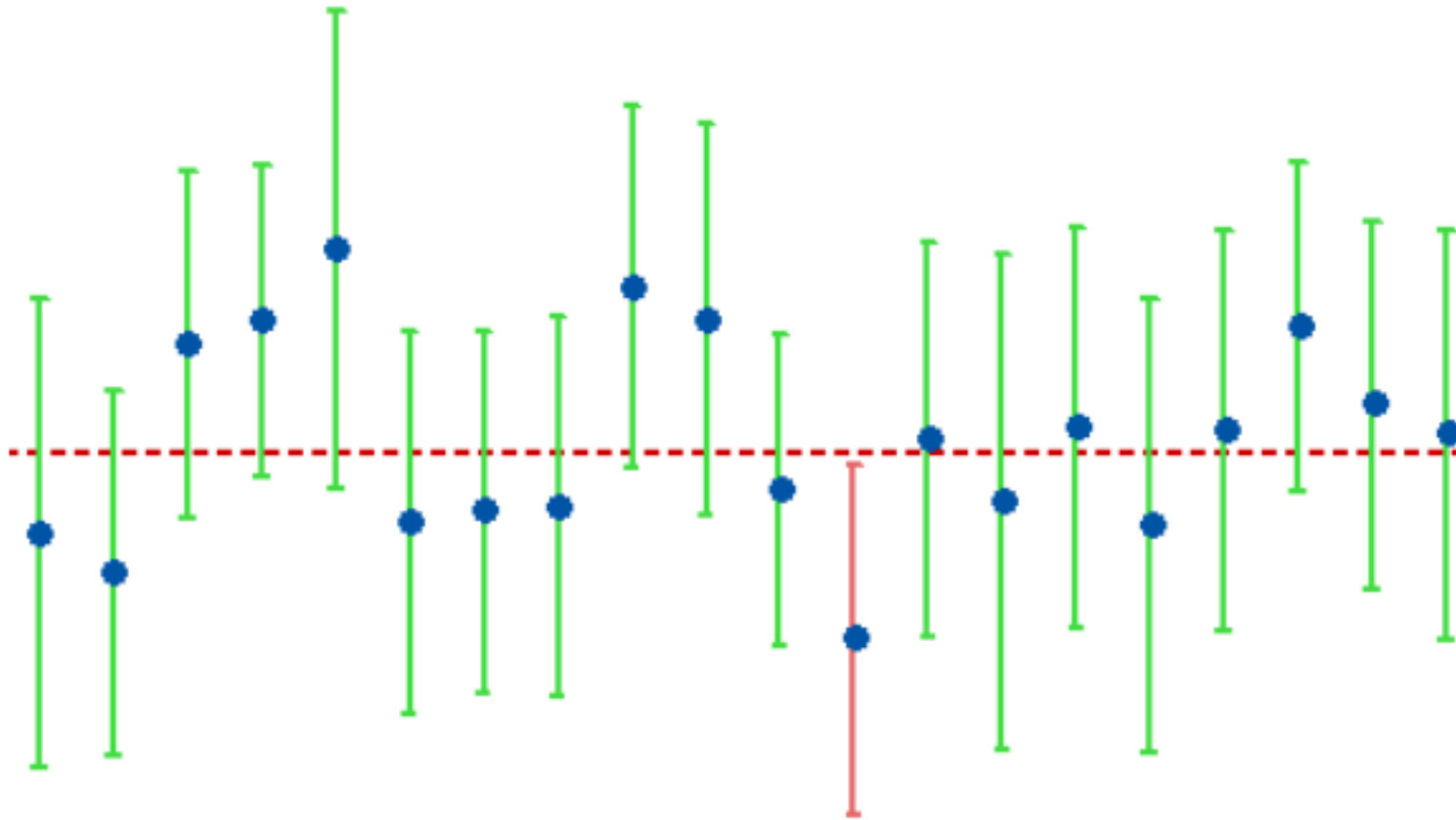
1. Konfidencia-intervallumok kiszámítása.
2. Hipotézis vizsgálat.

Konfidencia intervallum



Ötlet: Adjunk meg egy értéksávot, ami nagy valószínűséggel tartalmazza a tényleges β értékét!

Konfidencia intervallum: precízebben



Ha újra és újra kiszámítanánk a 95%-os konfidencia intervallumot, akkor az így keletkező intervallumok 95%-a tartalmazná a β valódi értékét.

Konfidencia intervallum

$$\text{Képlet: } [\hat{\beta} - t_{\beta} s_{\beta}, \hat{\beta} + t_{\beta} s_{\beta}]$$

Közepe a becsült β (azaz a $\hat{\beta}$).

Két tényező határozza meg az intervallum nagyságát: s_{β} ($\hat{\beta}$ szórása) és t_{β} (Student-féle eloszlásból jön).

A standard hiba (s_β)

A $\hat{\beta}$ is egy változó, értéke mintáról mintára változik. Értéke szóródik.

$\hat{\beta}$ szórásának képlete:

$$s_\beta = \sqrt{\frac{SSR}{(N - 2) \sum_{i=1}^N (X_i - \bar{X})^2}}$$

Korábbi tényezők hatása a standard hibára:

- A szóródást a maradéktagok okozzák. Ezért a standard hiba függ SSR-től. (SSR/(N-2) nem más, mint u varianciája.) Jobban szóródó maradéktagok – kevésbé pontos becslés.
- A nevezőben figyelembe vesszük X változékonyságát (a szummás tagban). Nagyobb változékonyság, kisebb hiba.
- Minél több megfigyelés, annál pontosabb a becslés. Standard hiba kisebb, minél nagyobb N.

A standard hiba (s_β)

A standard hiba képlete: $s_\beta = \sqrt{\frac{SSR}{(N-2) \sum_{i=1}^N (X_i - \bar{X})^2}}$

A konfidencia intervallum képlete: $(\hat{\beta} - t_\beta s_\beta, \hat{\beta} + t_\beta s_\beta)$

A standard hiba szerepe:

- Kisebb SSR \rightarrow kisebb s_β \rightarrow szűkebb konfidenciaintervallum

azaz a konfidenciaintervallum szélessége pozitívan függ az SSR nagyságától
(Nagyobb maradéktagok nagyobb bizonytalanságra utalnak.)

- Több megfigyelés (N) \rightarrow kisebb s_β \rightarrow szűkebb intervallum.
- Nagyobb X változékonyság \rightarrow kisebb s_β \rightarrow szűkebb intervallum.

A konfidencia intervallum

A konfidencia intervallum képlete: $(\hat{\beta} - t_{\beta} s_{\beta}, \hat{\beta} + t_{\beta} s_{\beta})$

A konfidencia intervallum nő s_{β} -ben és t_{β} -ben.

Bizonytalanság (azaz, az, hogy mintánk mennyire felel meg a valóságnak) növeli az intervallumot.

Bizonytalanságot növeli a kis elemszám, nagy hibatagok, és ha X nem változékony.

t_β

Egy bizonyos eloszlásból jön, részletek jövőre.

A t_β szerepe az, hogy a kívánt megbízhatósági szint függvényében felnagyítja a standard hiba hatását.

Ha 95% megbízhatósági szinttel dolgozunk, akkor t_β más, mint amikor a megbízhatósági szint 99%.

A standard hiba felnagyítása a megbízhatósági szint függvényében nem lineáris, sokkal bonyolultabb. De nekünk csak a megfelelő t_β értékeket kell használni a konfidencia-intervallum kiszámolásához. Illetve kiszámolja nekünk az Excel.

t_{β} **Table T** Critical Values of the t Distribution

df	One-Tail = .4 Two-Tail = .8	.25 .5	.1 .2	.05 .1	.025 .05	.01 .02	.005 .01	.0025 .005	.001 .002	.0005 .001
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.214	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

Source: From *Biometrika Tables for Statisticians*, Vol. 1, Third Edition, edited by E. S. Pearson and H. O. Hartley, 1966, p. 146.
 Reprinted by permission of the Biometrika Trustees.

A konfidencia intervallum - értelmezés

Leggyakoribb:

95%-os konfidenciaintervallum

„95% a valószínűsége, hogy az együttható valódi értéke az adott intervallumba esik”

„Ha 100 mintát vennénk, akkor 95 esetben a becsült együttható benne lenne az intervallumban”

Ha N nagy, 95%: $t=1.96$

Excel: megbízhatósági szint megadható

Példa 1 – munkanélküliség és GDP/fő

Mennyire lehetünk biztosak abban, hogy a munkanélküliség hatása negatív a GDP/fő-re?

	<i>Koefficiensek</i>	<i>Standard hiba</i>	<i>Alsó 95%</i>	<i>Felső 95%</i>
Tengelymetsz.	4506.504	827.728	2767.512	6245.496
munkanélküliség	-194.286	76.55845	-355.129	-33.4422

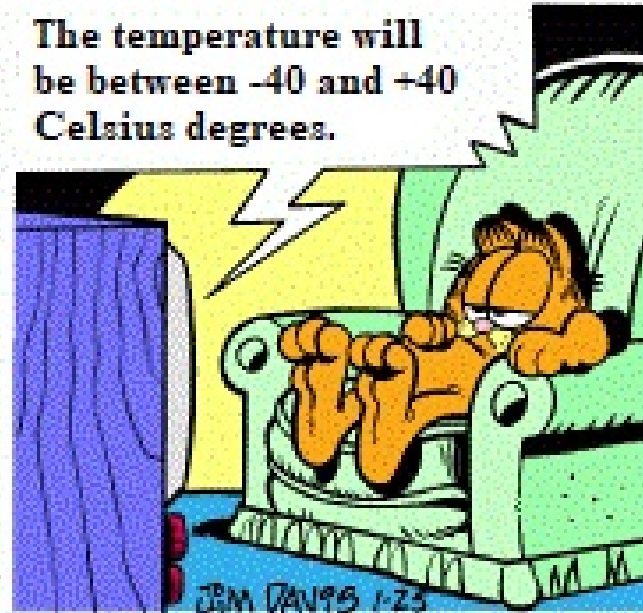
Példa 2 – hirdetés és értékesítés

Mennyire lehetünk biztosak abban, hogy az értékesítés hat az eladásokra?

	<i>Koefficiensek</i>	<i>Standard hiba</i>	<i>Alsó 95%</i>	<i>Felső 95%</i>
Tengelymetsz.	502.92	4.13	494.70	511.14
Hirdetés	0.22	0.08	0.06	0.37

Feladat - 2

Az órai excel fileban oldja meg a „Konfidencia intervallum” fülön található feladatokat!



Feladat - 3

Az órai excel fileban oldja meg a „Gyakorlás” fülön található feladatokat!

Házi feladat

A Moodle platformra feltöltve megtalálja a 6. házi feladatot, kövesse az ott megadott instrukciókat.

A számításai mellé írjon rövid értelmező szövegeket is.

Köszönöm a figyelmet





Corvinus

