Valószínűségszámítás és Statisztika

7. előadás

2017. november 5.

Statisztikai adatok grafikus ábrázolása

A grafikus ábrázolás célja az adatok alakulásának,egymáshoz viszonyított nagyságrendjének, arányainak érzékeltetése

A grafikus ábrázolás történhet

- vonalak,
- terület,
- térbeli alakzat és
- képszimbólumok

segítségével

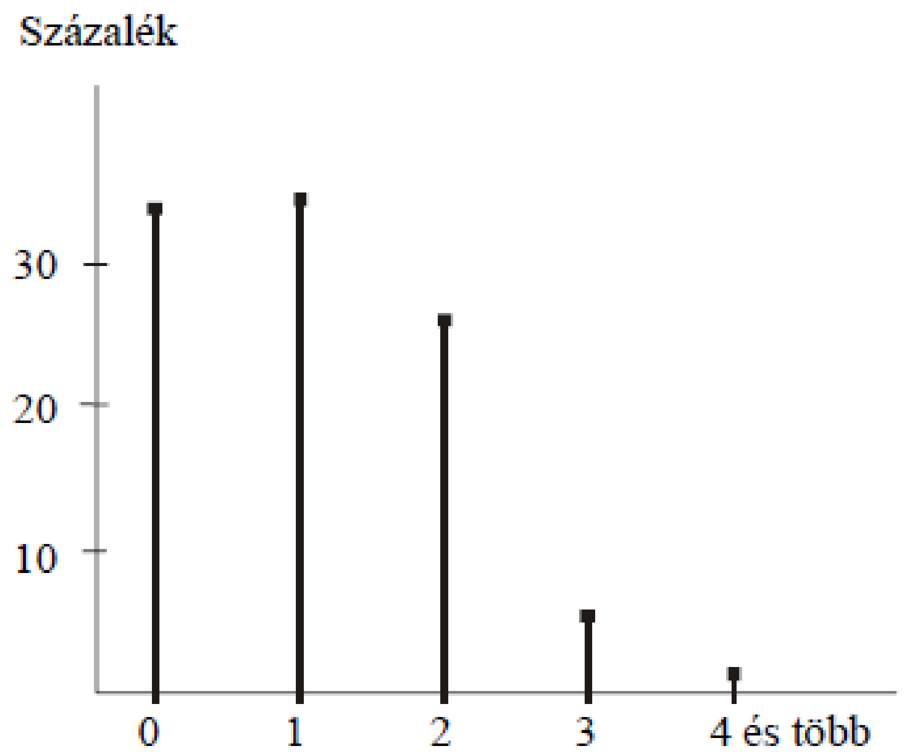
Ábratípusok

- Vonaldiagram
- Bot-diagram
- Poligon
- Hisztogram
- Oszlop- és sávdiagram
- Kördiagram
- Boxplot
- Terület-diagram
- Rúd-diagram, mértani testek
- Piktogram
- Kartogram

Vonaldiagram

Bot(pálcika)diagram

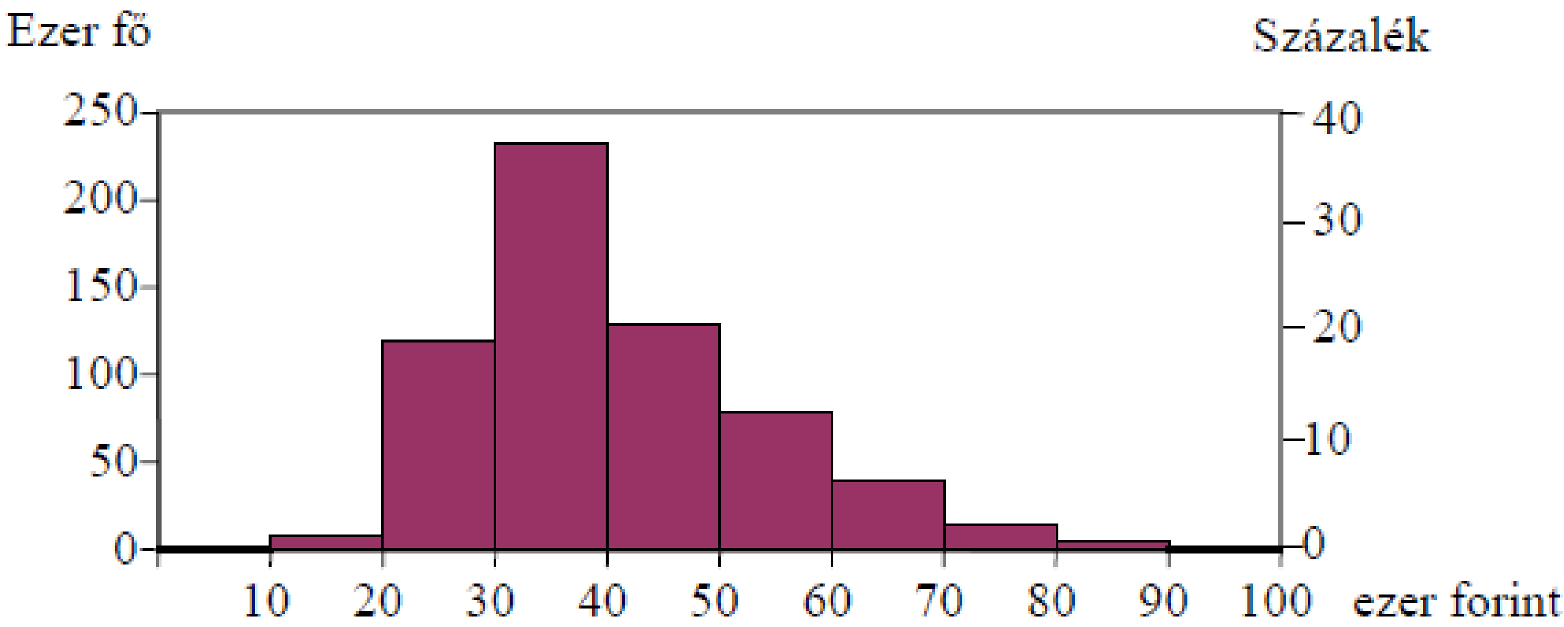
13. ábra. A családok gyermekszám szerinti megoszlása, 1996



Adatforrás: *Magyar statisztikai évkönyv, 1999* (2000). Központi Statisztikai Hivatal, Budapest.

Hisztogram

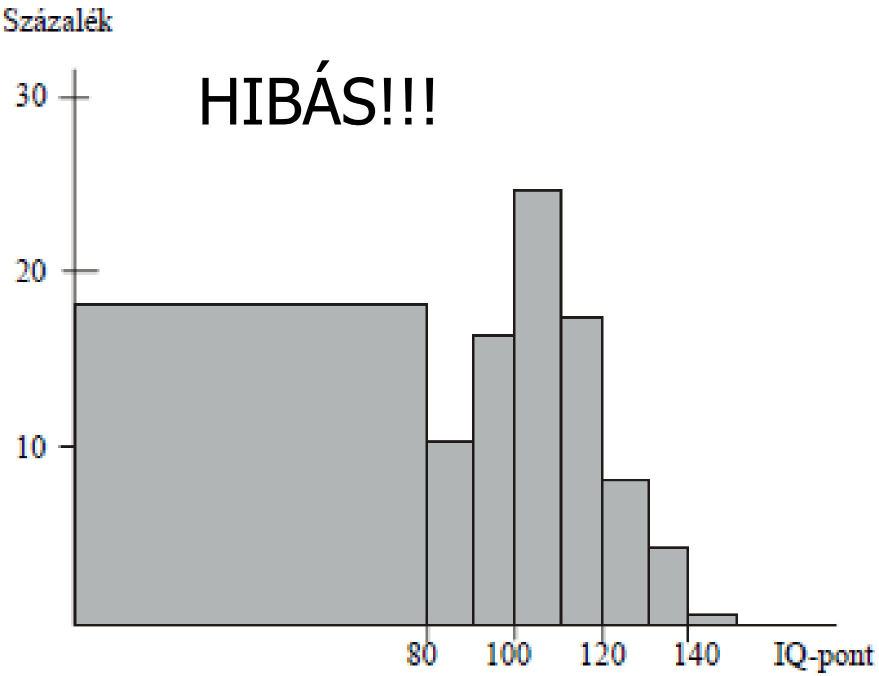
14. ábra. A férfi népesség megoszlása az öregségi nyugdíjak nagysága szerint
2000. január1.

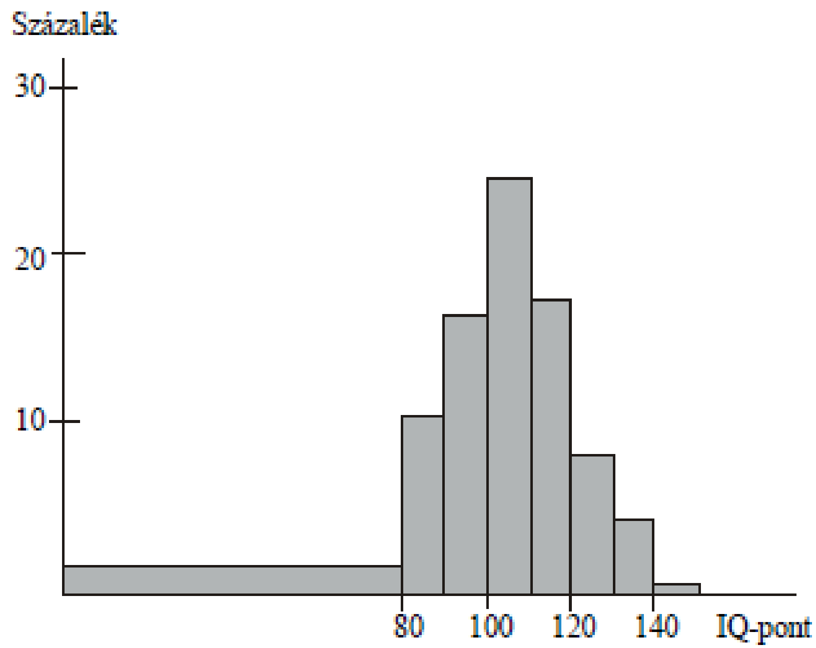


Adatforrás: Magyar statisztikai évkönyv, 1999 (2000). Központi Statisztikai Hivatal, Budapest.

Hisztogram (folyt.)

15. ábra. Egy népességcsoport megoszlása IQ-pontok szerint



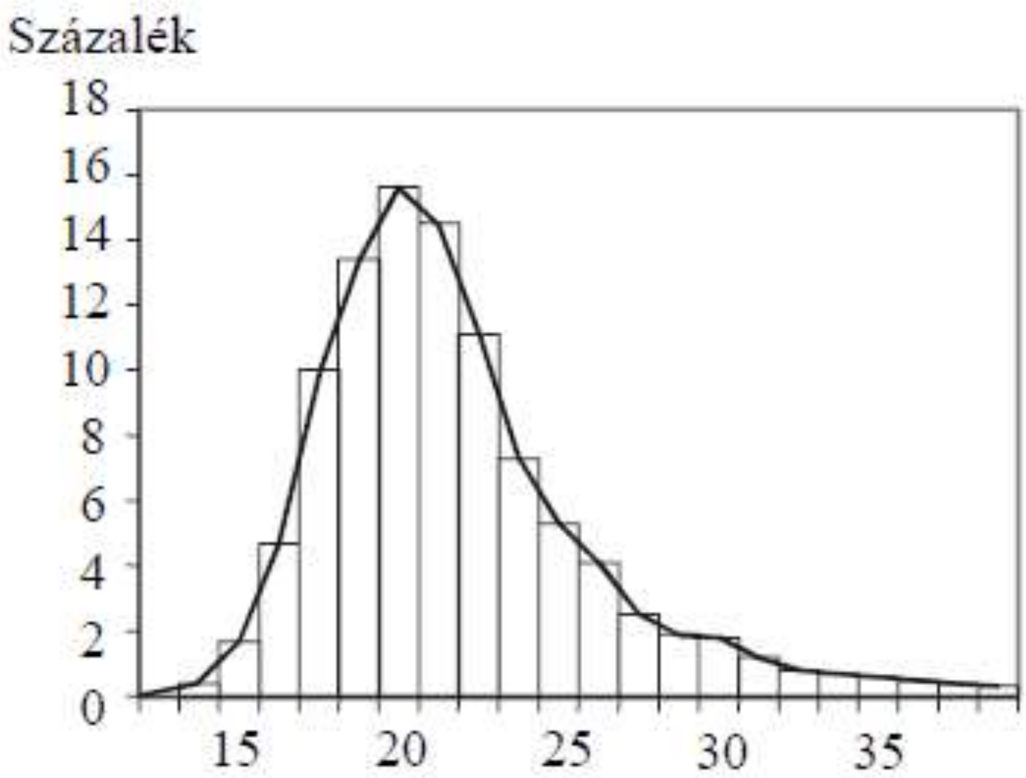


Egy népességcsoport IQ-pontok szerinti megoszlása

IQ-pont	Százalék	IQ-pont	Százalék	IQ-pont	Százalék
-80	18,2	101-110	24,7	131-140	4,1
81-90	10,8	111-120	17,3	141-150	0,5
91-100	16,2	121-130	8,2	*Összesen*	*100,0*

Poligon

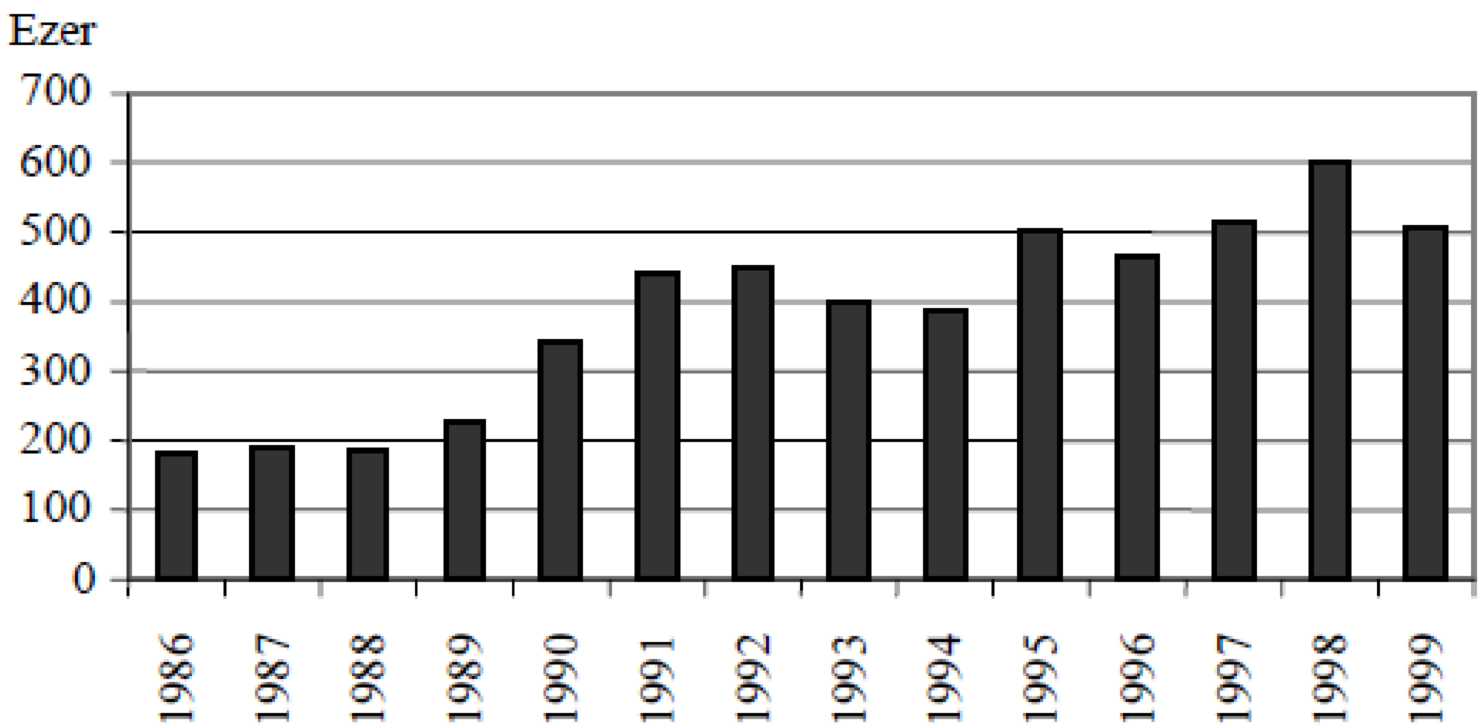
16. ábra. A sorköteles fiatalok BMI-index szerinti megoszlása



Forrás: Joubert – Gyenis; 2001.

Oszlopdiagram

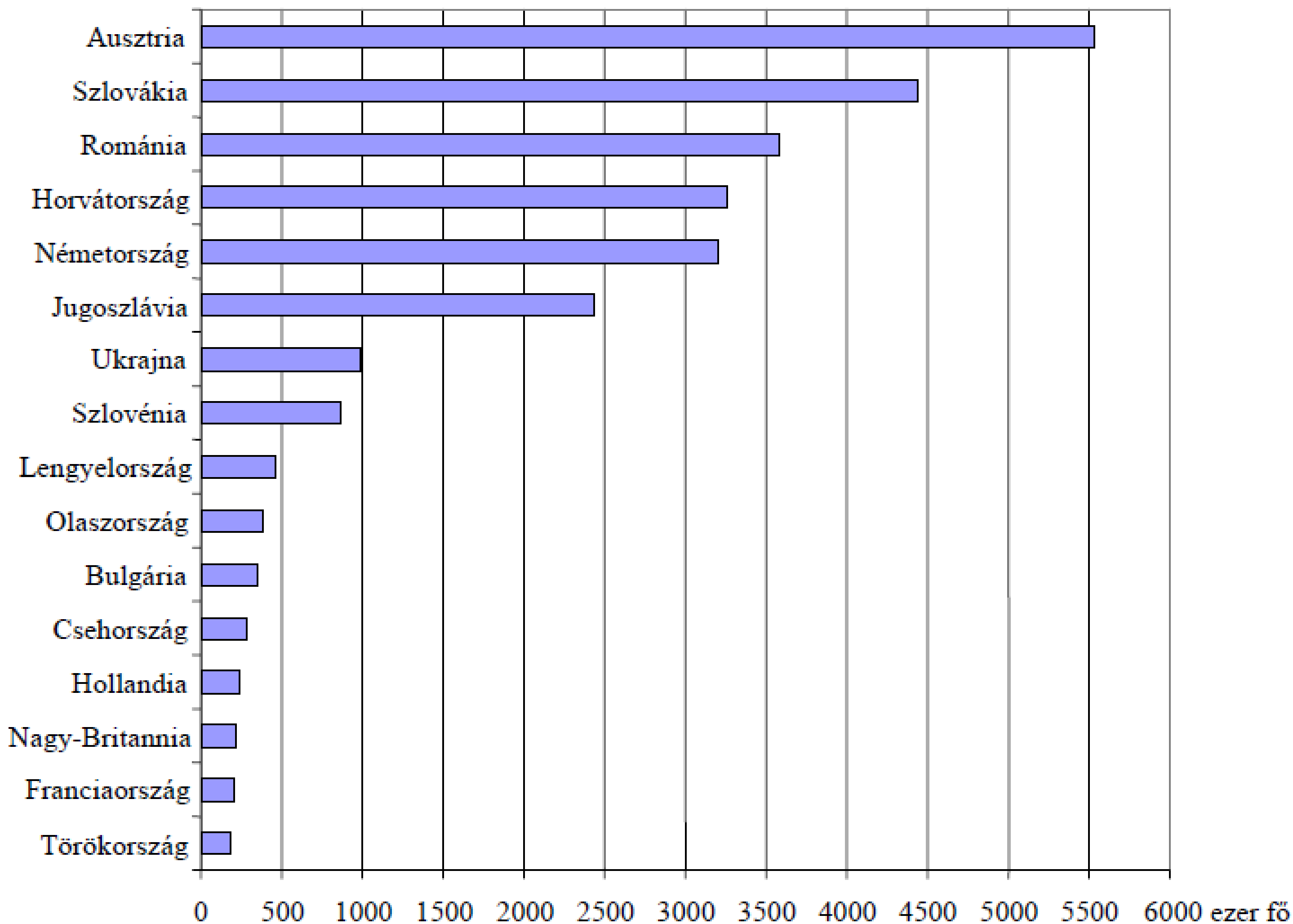
6. ábra. A bűncselekmények számának alakulása Magyarországon, 1986–1999



Adatforrás: *Magyar statisztikai évkönyv, 1999* (2000). Központi Statisztikai Hivatal, Budapest.

Sávdiagram

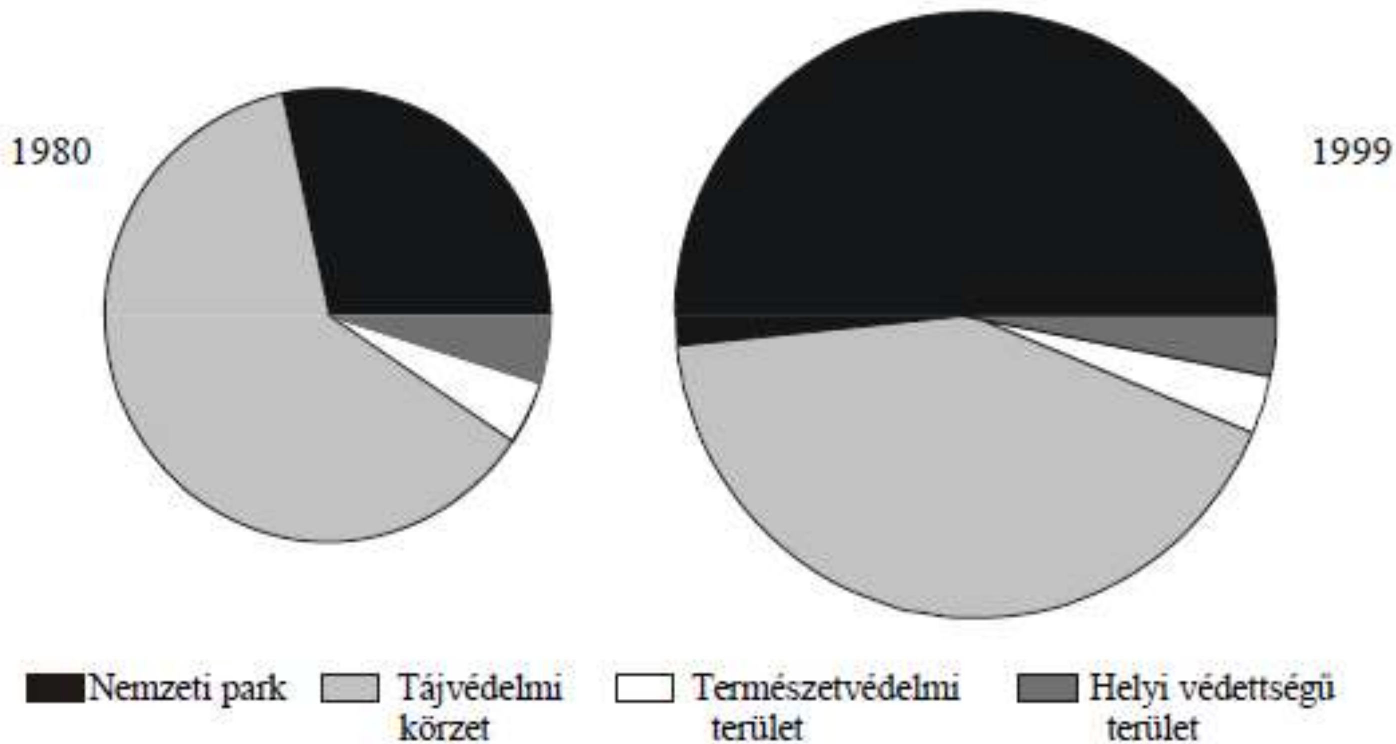
10. ábra. A Magyarországra néhány európai országból érkezett látogatók száma 1999-ben



Adatforrás: Magyar statisztikai évkönyv, 1999 (2000). Központi Statisztikai Hivatal, Budapest.

Kördiagram

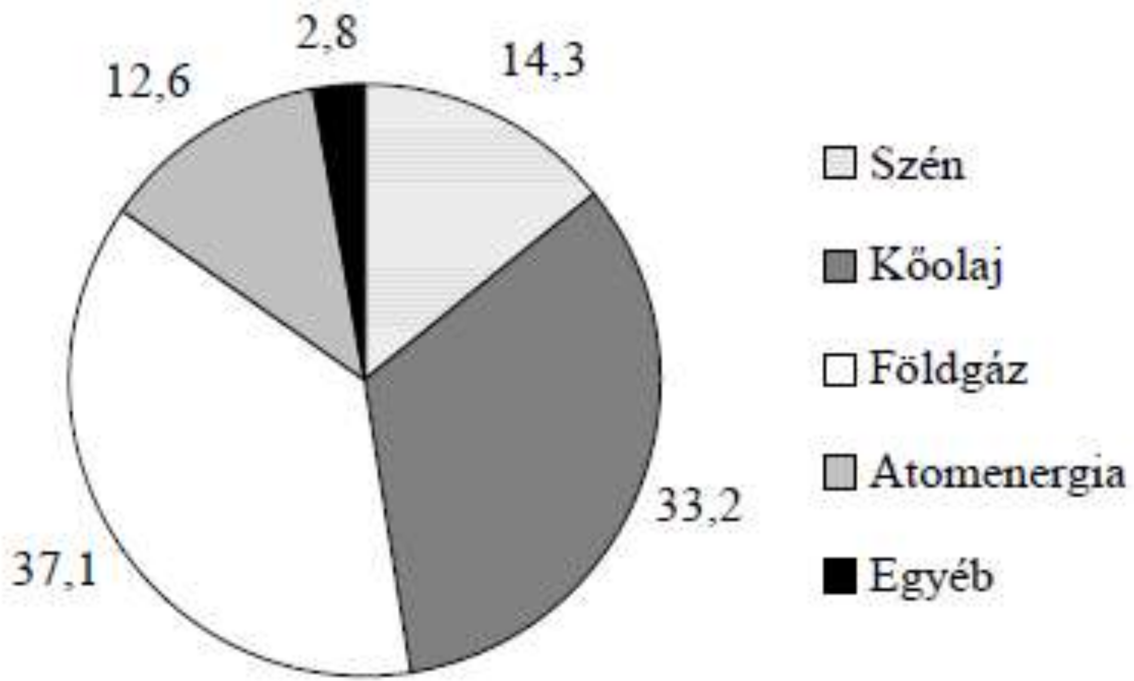
11. ábrák. Védett területek megoszlása Magyarországon 1980-ban és 1999-ben



Adatforrás: Magyar statisztikai évkönyv, 1999 (2000). Központi Statisztikai Hivatal, Budapest.

Kördiagram (folyt.)

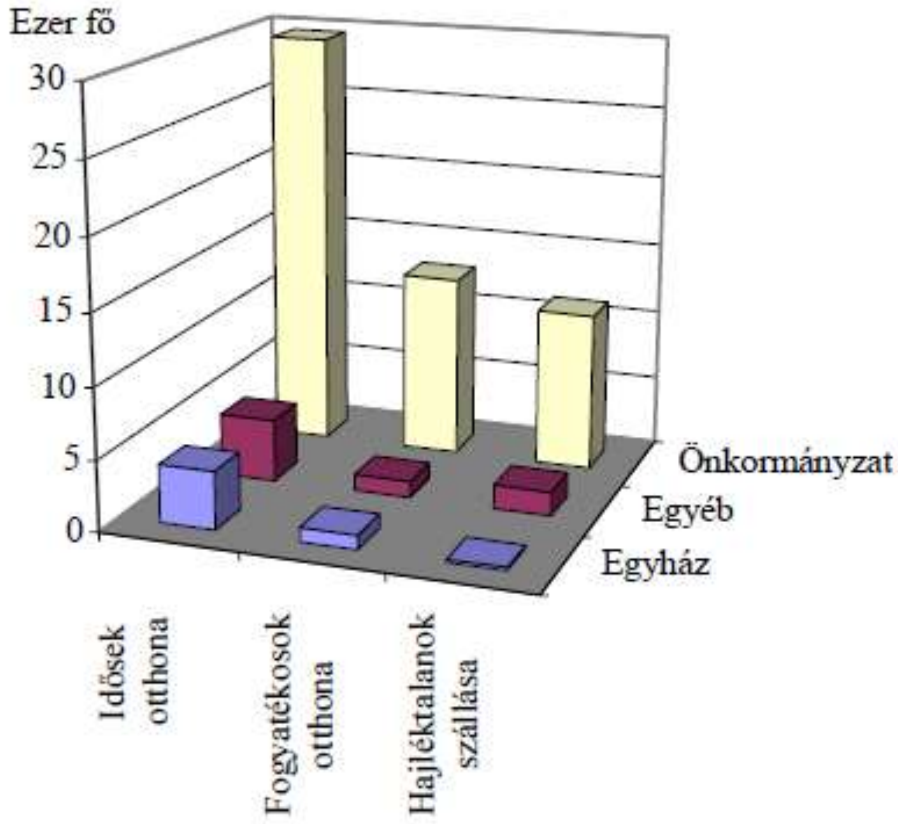
9. ábra. Az egyes energiahordozók százalékos aránya az összes felhasználásban, 1999



Adatforrás: *Magyar statisztikai évkönyv, 1999* (2000). Központi Statisztikai Hivatal, Budapest.

3D hisztogram

22. ábra. A szociális intézményekben gondozottak száma intézménytípus és fenntartó szerint, 1999



Adatforrás: Magyar statisztikai évkönyv, 1999 (2000). Központi Statisztikai Hivatal, Budapest.

Tapasztalati kvantilisek

- Elméleti kvantilis: abszolút folytonos, szigorúan monoton F esetén $q_z=F^{-1}(z)$
- Általában: $\inf\{x:F(x)>z\}$
- A tapasztalati eloszlás kvantilisei: tapasztalati kvantilisek. Esetleg lineáris interpolációval lehet pontosítani a becsléseinket.
- $z=1/2$: medián.
- $z=1/4$, $3/4$: kvartilisek

Medián

- A sorbarendezett minta középső eleme (ha páros sok eleme van: a két középső átlaga).
- Közelítés osztály-gyakoriságokra:

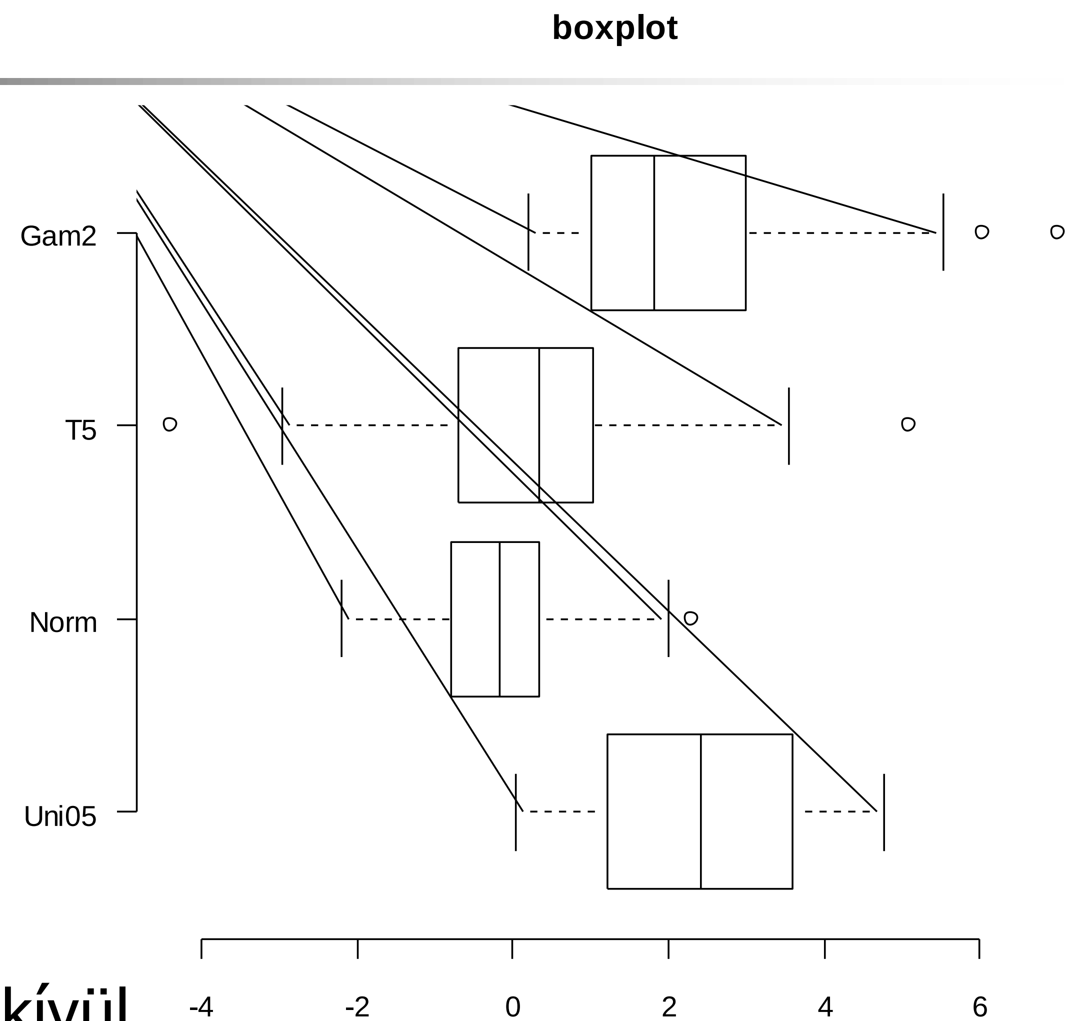
$$Me = x_l + \frac{\frac{n}{2} - f'_{me-1}}{f_{me}} h$$

  - x_l: a mediánt magában foglaló osztály alsó határa
  - f'_{me}: kumulált gyakoriság a mediánt megelőző osztályig bezárólag
  - f_{me} : a mediánt magában foglaló osztály gyakorisága
  - h: a mediánt magában foglaló osztály szélessége.
  - n: a minta elemszáma

Módusz

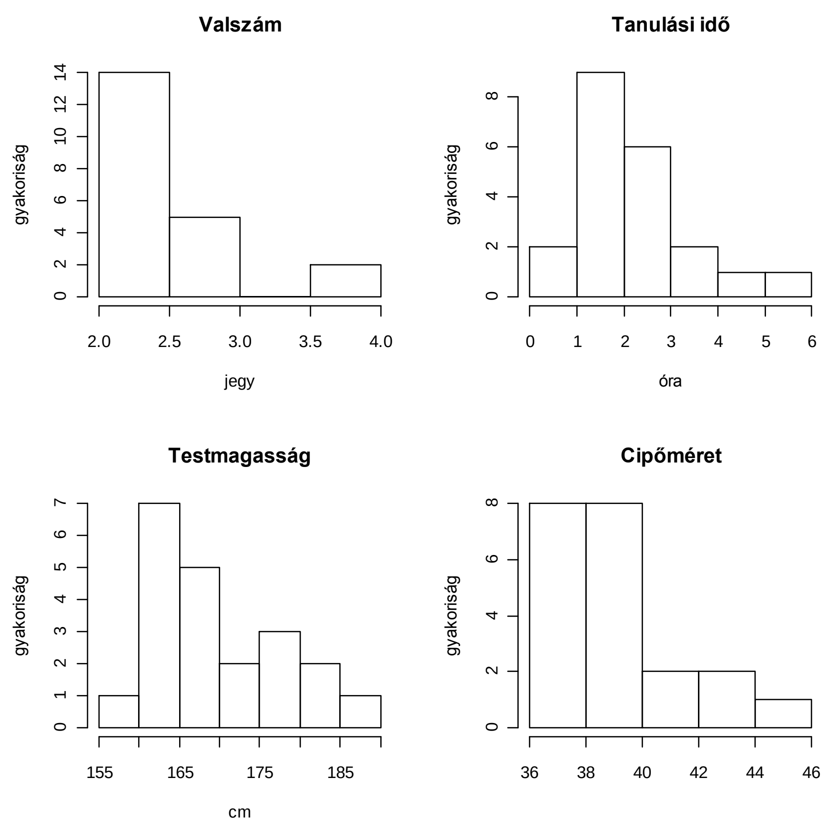
- A leggyakoribb (tipikus) érték.
- Az eloszlás lehet unimodális, bimodális vagy polimodális(egy-, két- vagy többmóduszú).
- Meghatározása: A gyakorisági poligon maximumhelye (a modális osztályköz középértéke).

Alapstatisztikák grafikus megjelenítése

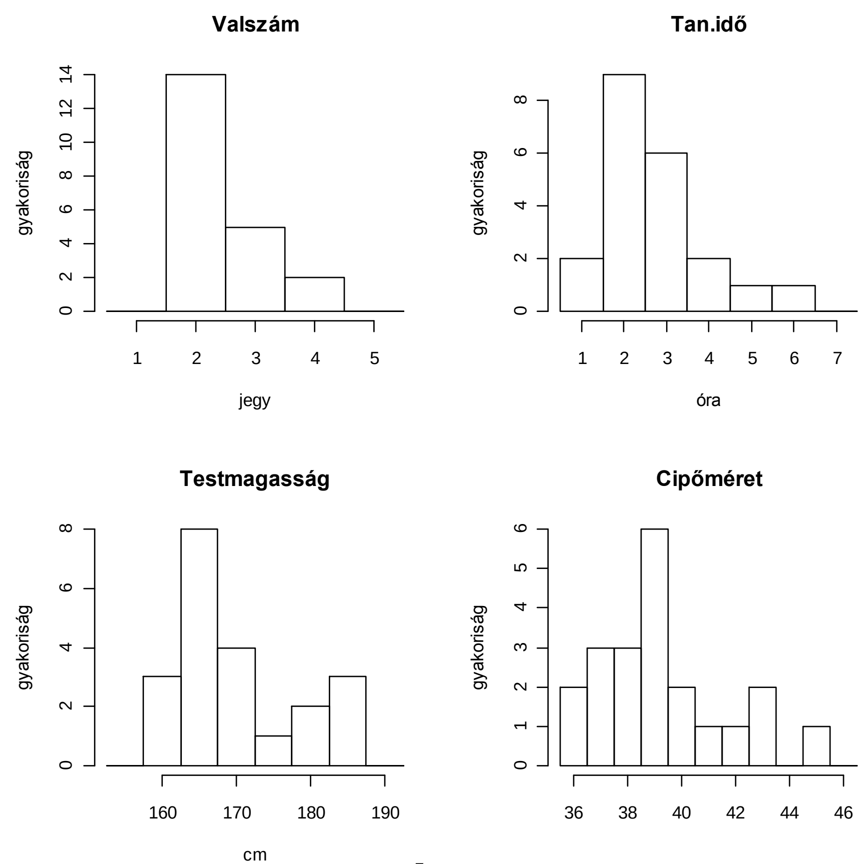


Az egyes dobozok az alsó kvartilistól a felső kvartilisig tartanak. Közép-vonal a medián. A vonalak a teljes terjedelmet felölelik, ha ez az egyes irányokban nem nagyobb a kvartilisek közötti különbség 1.5-szeresénél. Ha ezen kívül is vannak pontok, azokat külön-külön jeleníti meg.

Saját adataink

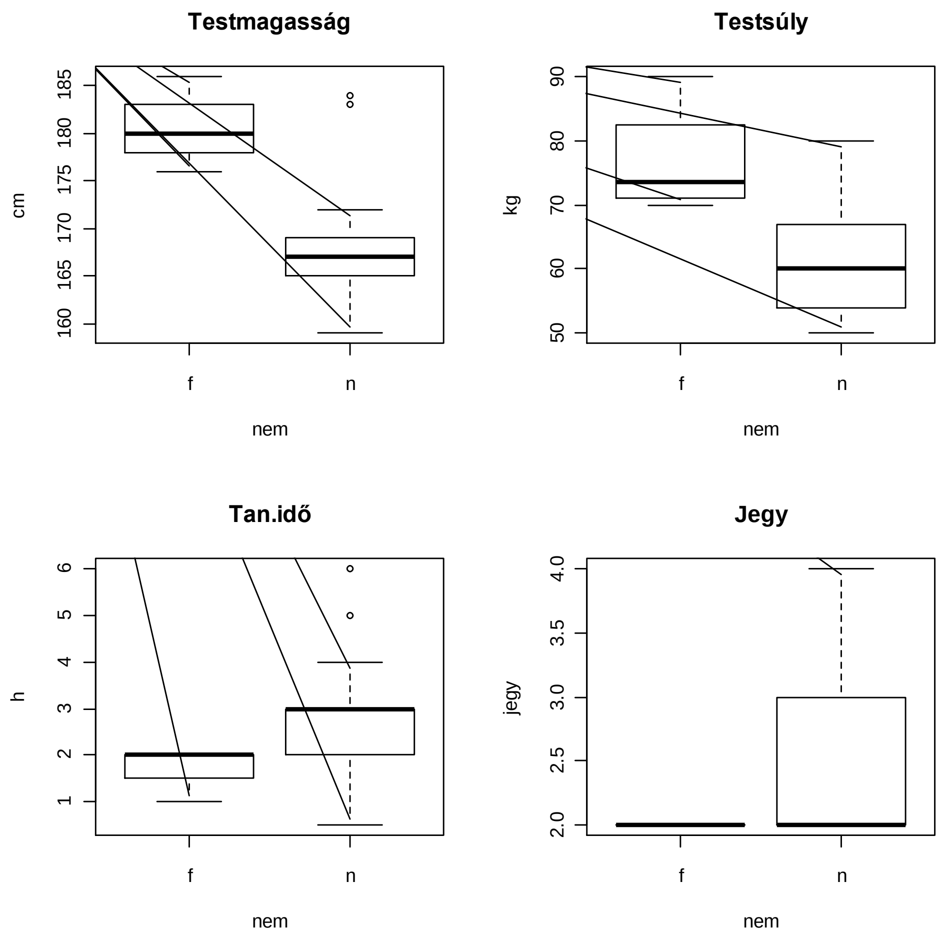


Alapbeállításokkal, rossz



Jó

Nemenkénti bontás



A későbbiekben térünk majd arra vissza, hogy mely különbségek szignifikánsak (lényegesek)

Táblázatok elemzése, viszonyszámok

- A számokat legjobban osztással tudjuk összehasonlítani. (Figyelem: különbség-, illetve összegképzés csak akkor értelmes, ha ez a halmazokra is értelmes).

A VISZONYSZÁMOK

A viszonyszám két egymással logikai kapcsolatban álló statisztikai adat hányadosa.

$$V = \frac{A}{B}$$

A viszonyszámok típusai

- **Megoszlási** viszonyszámok:
- **intenzitási** viszonyszámok:
  különböző fajta, rendszerint különböző mértékegységű adatokból számított
- **dinamikus** viszonyszámok

A megoszlási viszonyszám

- A sokaság egyes részeinek a sokaság egészéhez viszonyított arányát fejezi ki.
- Például: A csoporton belül a nők aránya 60%.

A dinamikus viszonyszám

Két időszak (időpont) adatának hányadosa.

Például: 2003-ban tízszer akkora volt az internethasználók száma, mint 1993-ban.

Intenzitási viszonyszám

Azt fejezi ki, hogy *az egyik mennyiségből mennyi jut a másik egy egységére*.

Például: Az 1 lakosra jutó napi tejfogyasztás 0,5 liter.

Az intenzitási viszonyszám típusai

- **Nyers viszonyszám:**

  Az összehasonlításban szereplő mennyiséget a teljes sokasághoz viszonyítjuk
  (pl. 1 családra jutó gyerekszám)

- **Tisztított viszonyszám:**

  Az összehasonlításban szereplő mennyiséget a sokaság azon részéhez viszonyítjuk, amellyel szorosabb logikai kapcsolatban van (ha pl. csak a gyerekes családokat vesszük figyelembe.)

Dinamikus viszonyszámok **Bázis- és láncviszonyszámok(indexek)**

- Mérőszám-sorozatok:
  - Bázisindex: idősor számait ugyanahhoz a bázisidőponthoz hasonlítjuk (egyszerű súlyozatlan index)

$$Bi_n = 100\frac{p_n}{p_0}$$

  - Láncindex: idősor egymás utáni számait hasonlítjuk egymáshoz

$$Li_n = 100\frac{p_n}{p_{n-1}}$$

Mérőszám-sorozatok, példa

- A táblázat harmadik oszlopa az éves áremelkedést mutatja.

Év	n	Bázisindex	Láncindex
1995	0	100	
1996	1	128,2	128,2
1997	2	158,5	123,6
1998	3	187,5	118,3
1999	4	214,3	114,3
2000	5	235,7	110
2001	6	258,8	109,8
2002	7	282,6	109,2
2003	8	297,6	105,3
2004	9	311,6	104,7

$$Bi_n = 100\frac{p_n}{p_0} = 100\prod_{j=1}^{n}\frac{p_j}{p_{j-1}} = 100\prod_{j=1}^{n}(Li_j/100)$$

Egy konkrét táblázat

- Magyarországi adatok
- Adjunk példát különböző arányszámokra!

	Népes-ség (Millió)	Születés-szám (ezer)	Autók száma (Millió)
1970	10,35	152	0,031
1980	10,70	149	0,230
1990	10,37	126	1
2000	10,17	97	1,96

Egy paradoxon

- 2 vállalkozás adatai
- Hol keresnek jobban az alkalmazottak?
- Adjuk meg mindkét cégnél az átlag-kereseteket!
- Tehát óvatosnak kell lennünk a kevert populációknál.

		Havi fizetés	Szám
B Bank	Nők	250	90
	Férfiak	350	10
G Gyár	Nők	200	10
	Férfiak	300	90

Indexszámok

- Indexszámok: Két hasonló ismérv adatát osztjuk el egymással.
- Egyszerű indexek
  - A kapcsolódó számok közvetlenül a táblázatból származnak, valódi mennyiségekről szólnak.
- (Összetett) indexszámok
  - A hasonlítandó értékek számok, amiket súlyozott átlagként kapunk meg.

Összetett indexek

- Összetett piacok időbeni változását jellemzik (átlagos ár- és mennyiségváltozás)
- A súlyok lényegesek, mert a termékek eltérő részt képviselnek a forgalomban.
- Két lehetőség:
  - 1. Laspeyres index : súlyok a bázisévből
  - 2. Paasche index : súlyok a beszámolási időszakból

Példa a saját tapasztalatunkból

- Heti kiadás, 2014: 1 mozijegy, 14 zsemle, 3 hamburger. Összár (érték): 800+14*50+3*300=2400 Ft
- Heti kiadás, 2015: 0,5 mozijegy, 10 zsemle, 7 hamburger. Összár (érték): 0,5*1200+10*60+7*300=3300 Ft, tehát egy 37,5%(=100*3300/2400)-os emelkedés. Ez egy értékindex, az ár- és mennyiségi változásokat nem különítettük el.

Folytatás: árösszehasonlítás

- Egy indexet keresünk, nemcsak egyszerű összehasonlításokat akarunk (50%,20%,0% az árunkénti árváltozás). A vásárlói kosár valódi összetételét kell figyelembe venni.
- A 2014-es mennyiségek alapján: 100*(1*1200+14*60+3*300)/(1*800+14*50+3*300)=100*2940/2400=122,5%, tehát 22,5%-os áremelkedés.
- A 2015-ös mennyiségek alapján: 100*(0,5*1200+10*60+7*300)/(0,5*800+10*50+7*300)=100*3300/3000=110%, tehát 10%-os volt az áremelkedés ebben az esetben. Alacsonyabb, mert kevesebbet fogyasztottunk a drágábbá vált árufélességekből.

Folytatás: mennyiségek összehasonlítása

- Figyelembe kell venni a fogyasztói kosár elemeinek árait.
- A 2014-es árak alapján: 100*(0,5*800+10*50+7*300)/(1*800+14*50+3*300)=100*3000/2400=120%, tehát 20%-kal nőtt a mennyiség ebben az esetben.
- A 2015-ös árak alapján : 100*(0,5*1200+10*60+7*300)/(1*1200+14*60+3*300)=100*3300/2940=112%, tehát 12%-kal nőtt a mennyiség ebben az esetben.
- Alacsonyabb, mert kevesebbet fogyasztottunk a megdrágult árukból.

Példa

- Utazási iroda adatai 2011-ből és 2012-ből.

	Meny-nyiség $q_{i,0}$	Meny-nyiség $q_{i,1}$	Átlagár (ezer Ft) $p_{i,0}$	Átlagár (ezer Ft) $p_{i,1}$
Év	2011	2012	2011	2012
Belföldi csoportos utak	200	223	52	52
Külföldi csoportos utak	132	128	208	240
Egyéni utak (menetjegyek)	188	192	74	80

Laspeyres-féle árindex

$$I_{0,1}^{P,L} = 100 \frac{\sum p_{i,1} q_{i,0}}{\sum p_{i,0} q_{i,0}}$$

I=100(52·200+240·132+80·188)/(52·200+208·132+74·188)=
=57120/51768=110,3

Tehát 10,3%-os áremelkedés volt ezen a piacon 2012-ben 2011-hez képest, ha a súlyok a bázisévből (2011) származnak.

Laspeyres-féle mennyiségi index

$$I_{0,1}^{M,L} = 100 \frac{\sum p_{i,0} q_{i,1}}{\sum p_{i,0} q_{i,0}}$$

I=100(52·223+208·128+74·192)/(52·200+208·132+74·188)=
=52428/51768=101,3

Tehát 1,3%-os mennyiségi növekedés volt ezen a piacon 2012-ben 2011-hez képest, ha a súlyok a bázisévből (2011) származnak.

Paasche –féle árindex

$$I_{0,1}^{P,P} = 100 \frac{\sum p_{i,1} q_{i,1}}{\sum p_{i,0} q_{i,1}}$$

I=100(52·223+240·128+80·192)/(52·223+208·128+74·192)=
=57676/52428=110,0

Tehát 10%-os áremelkedés volt ezen a piacon 2012-ben
2011-hez képest, ha a súlyok a beszámolási évből (2012) származnak.

Paasche-féle mennyiségi index

$$I_{0,1}^{M,P} = 100 \frac{\sum p_{i,1} q_{i,1}}{\sum p_{i,1} q_{i,0}}$$

I=100(52·223+240·128+80·192)/(52·200+240·132+80·188)=
=57676/57120=101,0

Tehát 1%-os mennyiségi növekedés volt ezen a piacon 2012-ben 2011-hez képest, ha a súlyok a beszámolási évből (2012) származnak.

Tulajdonságok

- Mindkét index a minimális és maximális (ár, illetve mennyiségi) változás-arány között helyezkedik el.
- Emelkedés $\rightarrow$ Index > 100%
- Ha minden ár/mennyiség ugyanúgy változik, akkor az index is ezt a hányadost adja.
- Időbeni változást nem tudjuk az indexek szorzatával megkapni, hanem csak indexsorokat tudunk kiértékelni. (egyszerűbb a Laspeyres-indexre).

Tulajdonságok

- Laspeyres index
  - Egyszerűbb számolni
  - Alkalmas indexsorok előállítására
  - Hajlamos a túlbecslésre
- Paasche index
  - Ár/mennyiség-aktuális
  - Hajlamos alulbecslésre
- Kompromisszum: Fischer-féle ideális index: geometriai közép a Paasche és a Laspeyres indexből.

$$I_{0,1}^{M,F} = \sqrt{I_{0,1}^{M,P} I_{0,1}^{M,L}} \qquad I_{0,1}^{P,F} = \sqrt{I_{0,1}^{P,P} I_{0,1}^{P,L}}$$

Értékindex

$$I_{0,1} = 100 \frac{\sum p_{i,1} q_{i,1}}{\sum p_{i,0} q_{i,0}}$$

I=100(52·223+240·128+80·192)/(52·200+208·132+74·188)=
=57676/51768=111,4

Azt jelenti, hogy a piacon forgalmazott összértékben 11,4%-os növekedés volt
megfigyelhető 2012-ben 2011-hez képest.

Felhasznált diák, anyagok

- Zempléni András: Leíró és matematikai statisztika
- ÁVF Leíró statisztika. Statisztikai alapismeretek http://www.avf.hu/tanarok/lipecz/avf-statisztika/stat-leiro/leiro-dia/leiro-01-alapfogalmak.ppt
- HUNYADI LÁSZLÓ: GRAFIKUS ÁBRÁZOLÁS A STATISZTIKÁBAN, Statisztikai Szemle, 80. évfolyam, 2002. 1. szám