STATISZTIKA

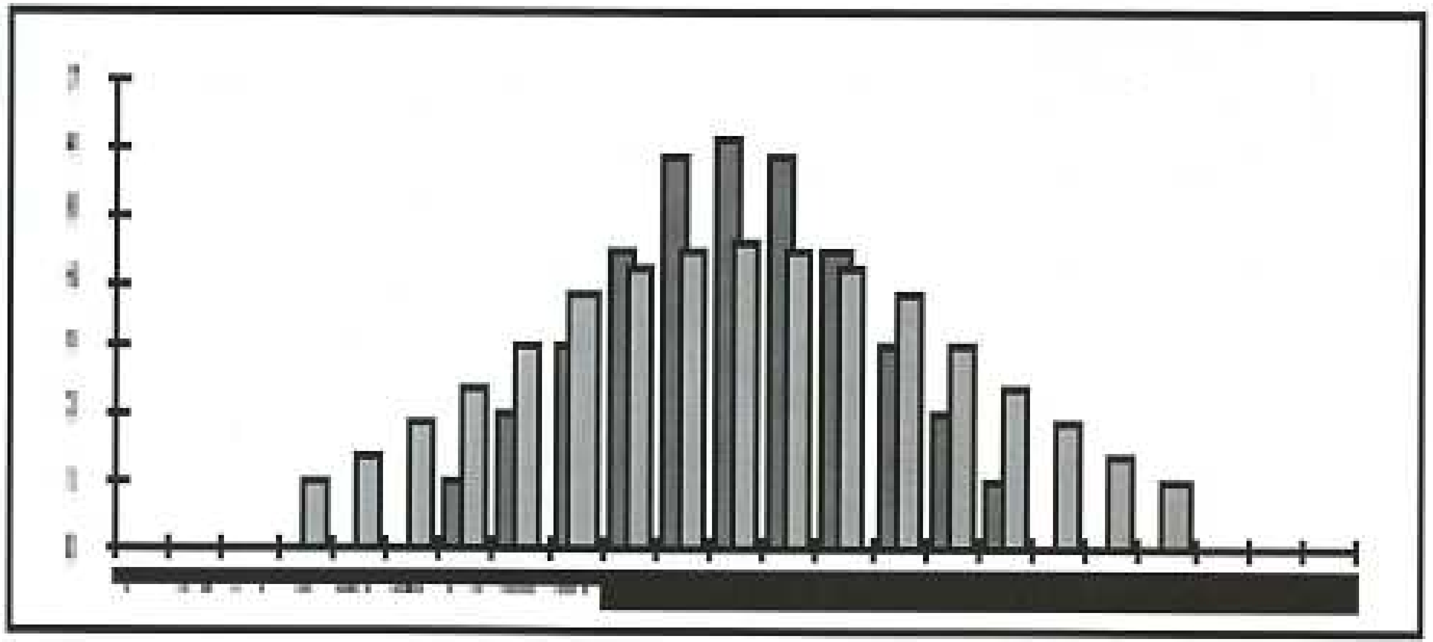
Mennyiségi ismérv szerinti elemzés (2)

Szóródás

2) Szóródás

Az értékek különbözőségét, változékonyságát nevezzük szóródásnak.

Az értékek különbözősége egyrészt az értékek egymástól való különbözőségén, másrészt valamely középértéktől való eltérésében fejezhető ki.



Szóródási mérőszámok

A legfontosabb szóródási mérőszámok:

1. Terjedelem, **R** (vagy **IQR**)
2. Átlagos eltérés, **δ**
3. Szórás, **б** (vagy **s** – minta esetén)
4. Relatív szórás, **V**
5. (Átlagos különbség, **G**)

Szélsőértékek és terjedelem típusú egyenlőtlenségi mutatók

- Maximum
  - Az adatsor legnagyobb értéke (x_{max})
- Minimum
  - Az adatsor legkisebb értéke (x_{min})
- Alapja a terjedelem típusú egyenlőtlenségi mutatóknak
  - Terjedelem (szóródás terjedelme) $P = x_{max} - x_{min}$
  - Terjedelem-arány (adatsor terjedelmének aránya) $K = \frac{x_{max}}{x_{min}}$
  - Relatív terjedelem $Q = \frac{x_{max} - x_{min}}{\bar{x}}$

Szélsőértékek és terjedelem típusú egyenlőtlenségi mutatók

- **Interkvartilis terjedelem** $\mathbf{IQR = Q_3 – Q_1}$

  Kifejezi, hogy mekkora értékkörben ingadozik az ismérvértékek középső 50%-a.

- **Interdecilis terjedelem** $\mathbf{IDR = D_9 – D_1}$

  Kifejezi, hogy mekkora értékkörben ingadozik az ismérvértékek középső 80%-a.

Szóródási mérőszámok

2) Átlagos eltérés: az átlagtól vett eltérések számtani átlaga.

Azt mutatja, hogy az ismérvértékek átlagosan mennyivel térnek el a számtani átlagtól.

Mértékegysége megegyezik az alapadatok mértékegységével, számítása:

Egyszerű:

$$\delta = \frac{\sum_{i=1}^{n} |d_i|}{n}$$

Súlyozott:

$$\delta = \frac{\sum_{i=1}^{k} f_i \cdot |d_i|}{\sum_{i=1}^{k} f_i}$$

$$d_i = x_i - \bar{x}$$

Szórás-típusú egyenlőtlenségi mutatók

- **Nem fajlagos (abszolút) mutató (x_i):** súlyozatlan szórás;
- **Fajlagos mutató (y_i):** súlyozott szórás;
- A valódi egyenlőtlenségeket a **relatív szórás**sal mérhetjük;
  - **Nem fajlagos mutató:** súlyozatlan relatív szórás (súlyozatlan szórás az átlag %-ában);
  - **Fajlagos mutató:** súlyozott relatív szórás (súlyozott szórás az átlag %-ában);

Súlyozatlan abszolút szórás: nem fajlagos mutatók esetében

- Az adott adatsor értékei (x_i) átlagtól való négyzetes eltérései átlagából vont négyzetgyök;
- **Képlete:**

  x_i = abszolút mutató az *i* régióban;

  n = elemszám;

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

- **Értékkészlete:** $0 \leq \sigma \leq \infty$

  Minél nagyobb az értéke, annál jobban szóródik x_i az átlag körül;

- **Mértékegysége:** mint az eredeti értékek (x_i) mértékegysége;

Súlyozatlan relatív szórás: nem fajlagos mutatók esetében

- A valódi egyenlőtlenségeket a relatív szórással mérhetjük;
- **Relatív szórás:** abszolút mutatók esetében;
- **Képlete:**

  σ = x_i adatsor szórása

  $\bar{x}$ = x_i adatsor átlaga

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

- **Kiszámítása:**

  a szórás értékeket elosztjuk az átlaggal és megszorozzuk 100-zal (a szórás értékeit az átlag %-ában fejezzük ki)

- **Értékkészlete:** $0 \leq v \leq \infty$;

  Minél nagyobb az értéke, annál jobban szóródik x_i az átlag körül;

- **Mértékegysége:** % ;

Súlyozott szórás: fajlagos mutatók esetében

- Fajlagos mutatók (y_i) esetében;
- Az adott adatsor értékei (y_i) átlagtól való négyzetes eltérései súlyozott átlagából vont négyzetgyök;
- Képlete:

  y_i = fajlagos mutató *i* régióban;

  f_i = súly;

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

- Értékkészlete: $0 \leq \sigma \leq \infty$;

  Minél nagyobb az értéke, annál jobban szóródik y_i az átlag körül;
- **Mértékegysége:** mint az eredeti értékek (y_i) mértékegysége;

Súlyozott relatív szórás: fajlagos mutatók esetében

- A valódi egyenlőtlenségeket a relatív szórással mérhetjük;

  Fajlagos mutatók esetében: súlyozott relatív szórással;

- **Képlete:**

  σ = y_i adatsor súlyozott szórása;

  $\bar{y}$ = y_i adatsor súlyozott átlaga;

$$v = \frac{\sigma}{\bar{y}} * 100$$

- **Kiszámítása:**

  A súlyozott szórást elosztjuk a súlyozott átlaggal és megszorozzuk 100-zal (a súlyozott szórást a súlyozott átlag %-ában fejezzük ki);

- **Értékkészlete:** $0 \leq v \leq \infty$;

  Minél nagyobb az értéke, annál jobban szóródik y_i az átlag körül;

- **Mértékegysége:** % ;

Példa: A szóródás jellemzői

- **Adatok:** 1 2 4 1; rendezve: 1 1 2 4
- **Terjedelem:** max-min=4-1=3
- **Kvartilisek:**
- **Standard eltérés:**

Percentiles

	Percentiles		
	25	50	75
Weighted Average(Definition 1)	1.0000	1.5000	3.5000
Tukey's Hinges	1.0000	1.5000	3.0000

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	1-2=-1	1
1	1-2=-1	1
2	2-2=0	0
4	4-2=2	4
Összeg	0	6

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{6}{3}} = \sqrt{2} = 1.414$$

Az átlag hibája (standard error, SE);

- Azt fejezi ki, hogy az átlag, amit a mintából számoltunk, mennyire megbízható;
- Ha többször is meg tudnánk ismételni a mérést (végtelen sok mérési sorozatunk lenne), akkor a mindegyik mérési sorozatra kapott átlag szóródását mutatja a populáció átlaga körül;
- **Számítása:** Standard error (SE): $SE=SD/\sqrt{n}$;

Standard deviáció, vagy standard error?

- **Standard deviáció (SD):** a minta szórása, azaz a mintaadatok szóródása az átlag körül;
  Normális eloszlás esetén az átlag ±2SD-n belül van az adatok kb. 95%-a;
- **Standard error (SE=SD/√n):** az átlag megbízhatósága, a mintaátlag szóródása az (ismeretlen) populáció átlag körül;
  Normális eloszlás esetén az átlag ± 2SE-n belül van az igazi átlag kb. 95%-os valószínűséggel;

A szórás tulajdonságai (1)

- Ha minden x értékhez ugyanazt a konstans számot hozzáadjuk (x+a), a szórás változatlan marad;
- Ha minden x értéket ugyanazzal a *k* konstans számmal megszorozzuk (kx), a szórás is *k*-szorosára változik;
- Az eltérésnégyzet-összeg az ***átlagtól való eltérésekkel*** számolva a ***legkisebb;***
- A **szórásnégyzet** felírható a **négyzetes átlag** és a **számtani átlag négyzetének** a különbségeként, azaz:

$$\sigma^2 = \bar{x}_q^2 - \bar{x}^2$$

A szórás tulajdonságai (2)

- A sokaságot jellemző **teljes szórásnégyzet** (variancia) megegyezik a rész-sokaságok **külső és belső szórásnégyzetének** összegével;

$$\sigma^2 = \sigma_B^2 + \sigma_K^2$$

- Értéke 0, ha x = konstans;
- Értékhatára

$$0 \leq \sigma \leq \bar{x}\sqrt{N-1}$$

Példa a szórás tulajdonságaira

x_i	$d_i = x_i - \bar{x}$	d_i^2	y_i	$d_i = y_i - \bar{y}$
100	-100	10000	150	-100
150	-50	2500	200	-50
210	+10	100	260	+10
240	+40	1600	290	+40
300	+100	10000	350	+100
Σ *1000*	*0*	*24200*	*1250*	*0*
$\bar{x}$ **200**			$\bar{y}$ **250**	
		σ^2**=4840**		σ^2**=4840**
		σ**=69,6**		σ**=69,6**

Példa a szórás tulajdonságaira

	x_i	$d_i = x_i - \bar{x}$	d_i^2		y_i	$d_i = y_i - \bar{y}$	d_i^2
	100	-100	10000		110	-110	12100
	150	-50	2500		165	-55	3025
	210	+10	100		231	+11	121
	240	+40	1600		264	+44	1936
	300	+100	10000		330	+110	12100
Σ	*1000*	*0*	*24200*		*1100*		*29282*
$\bar{x}$	**200**			$\bar{y}$	**220**		
			σ^2=4840				**σ^2=5856,4**
			σ=69,6				**σ=76,52**

Példa a szóródásra és szórásra (1)

Vízfogyasztás (m^3) (osztályközép: x_i)	Lakások száma (f_i)	f'
– 15	5	5
15 – 25	17	22
25 – 35	15	37
35 – 45	8	45
45 –	5	50
Összesen	50	-

$$\sigma = \sqrt{\frac{5(10-28,2)^2 + 17(20-28,2)^2 + \ldots + 5(50-28,2)^2}{50}}$$

$$\sigma = 11,26\, m^3$$

$$V = \frac{\sigma}{\bar{x}} = \frac{11,26}{28,2} = 0,4$$

Példa a szóródásra és szórásra (2)

10 14 15 23 58

Öt ember vagyona, mFt-ban.

Határozzuk meg a

a) Átlagot

b) Átlagos abszolút eltérést

c) Szórásnégyzetet, szórást, relatív szórást

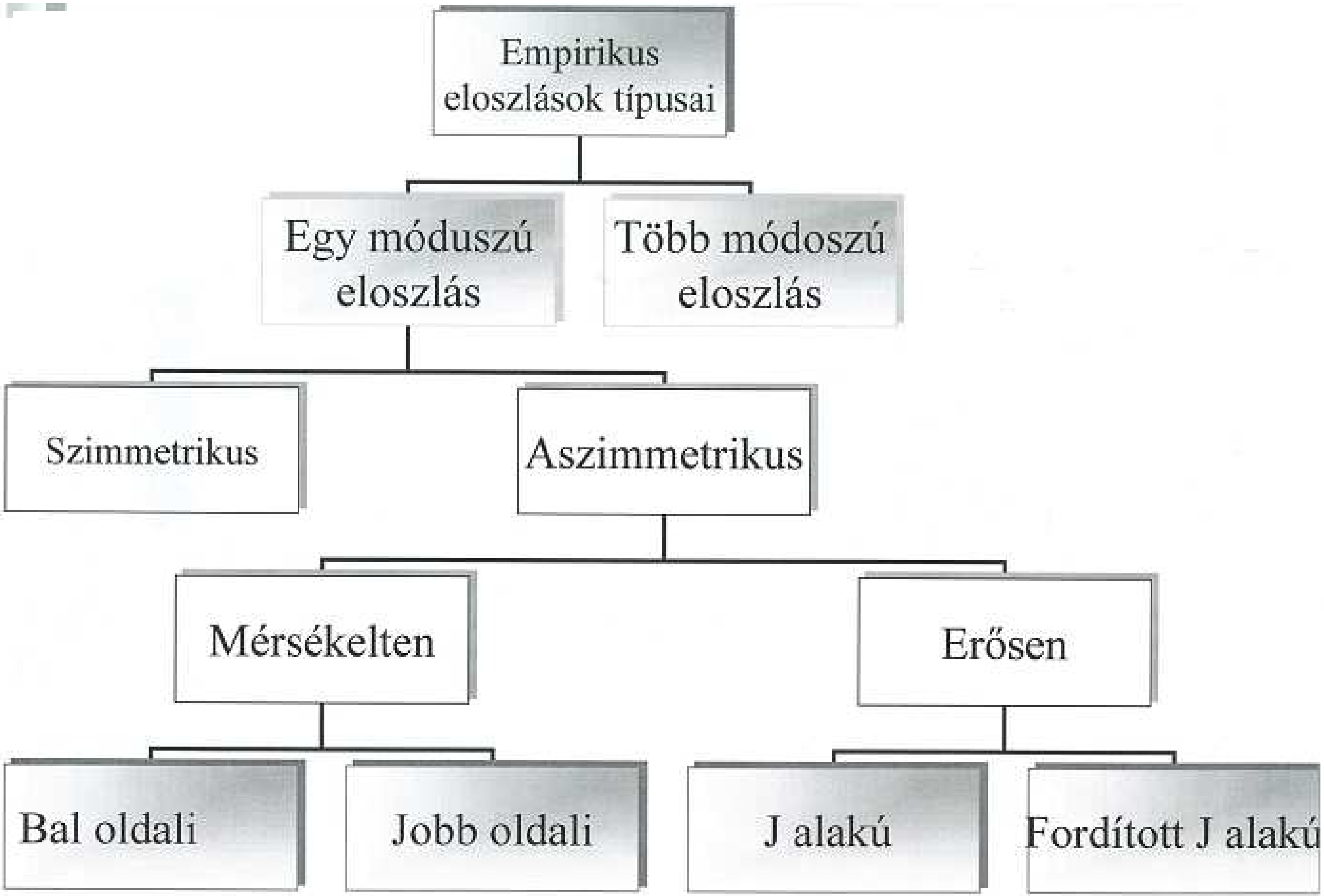
Értelmezzük az eredményeket!

Szóródási mérőszámok

5) **Átlagos különbség, G (Gini-féle mutató)** az ismérvértékek egymástól mért abszolút eltéréseinek a számtani átlaga. (Leginkább a koncentráció vizsgálatánál alkalmazható.)

$$G = \frac{1}{n^2} \cdot \sum_{j=1}^{n} \sum_{i=1}^{n} \left| x_i - x_j \right|$$

$$G = \frac{1}{n^2} \cdot \sum_{j=1}^{k} \sum_{i=1}^{k} f_i \cdot f_j \cdot \left| x_i - x_j \right|$$

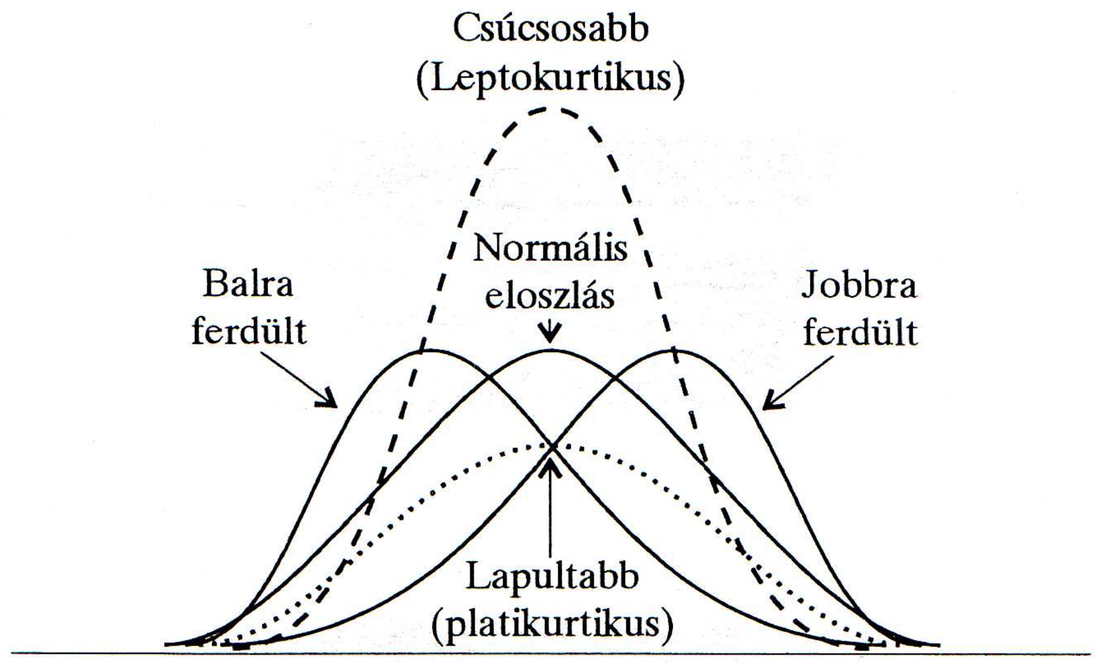

Empirikus eloszlások típusai
Egy móduszú eloszlás
Több módoszú eloszlás
Szimmetrikus
Aszimmetrikus
Mérsékelten
Erősen
Bal oldali
Jobb oldali
J alakú
Fordított J alakú

3) Alakmutatók

arra szolgálnak, hogy tömör számszerű formában jellemezzék, hogy milyen tekintetben és milyen mértékben tér el az adott eloszlás a normális eloszlás gyakorisági görbéjéből.

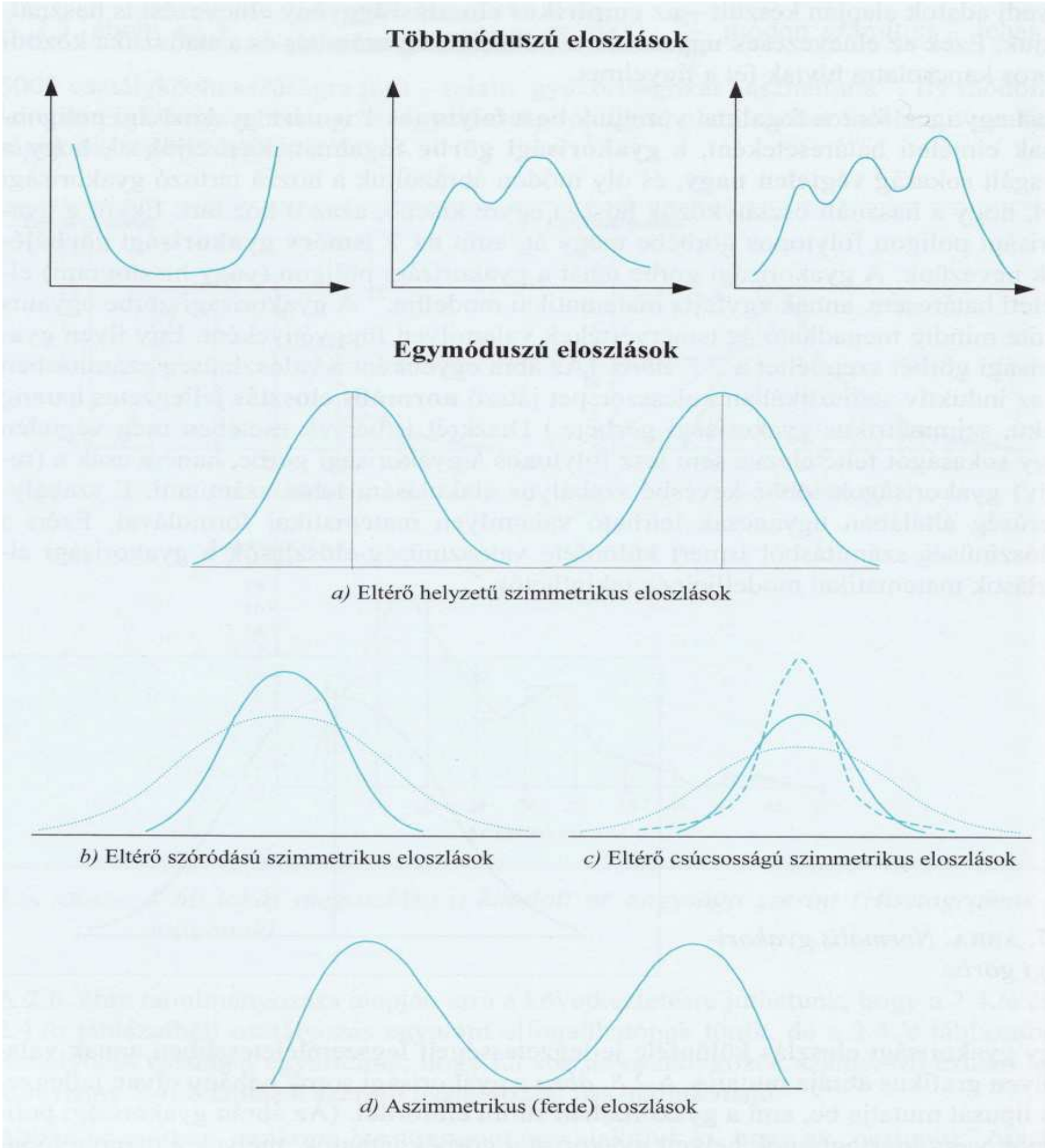
Mértékegység nélküli mutatók.

Alakmutatók, és helyzetmutatók



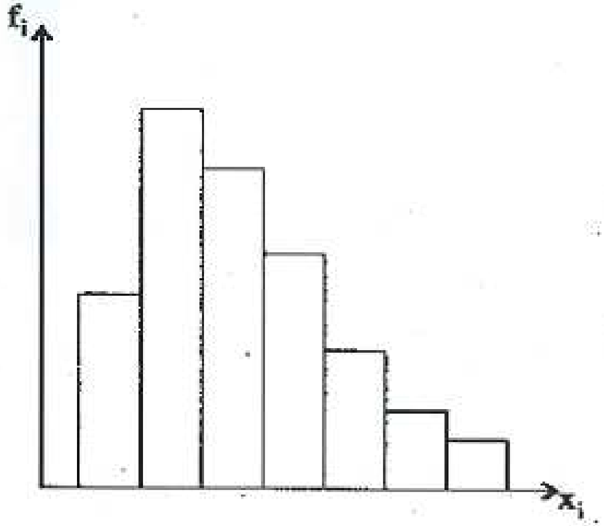
Az egymóduszú gyakorisági eloszlások lehetséges eltérései a normális gyakorisági görbétől

Eltérő jellegzetességű gyakorisági eloszlások

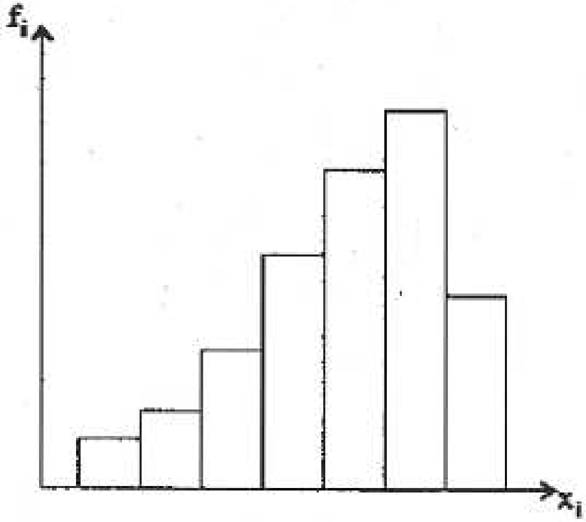


Aszimmetrikus eloszlások

Bal oldali aszimmetria **Jobb oldali aszimmetria**



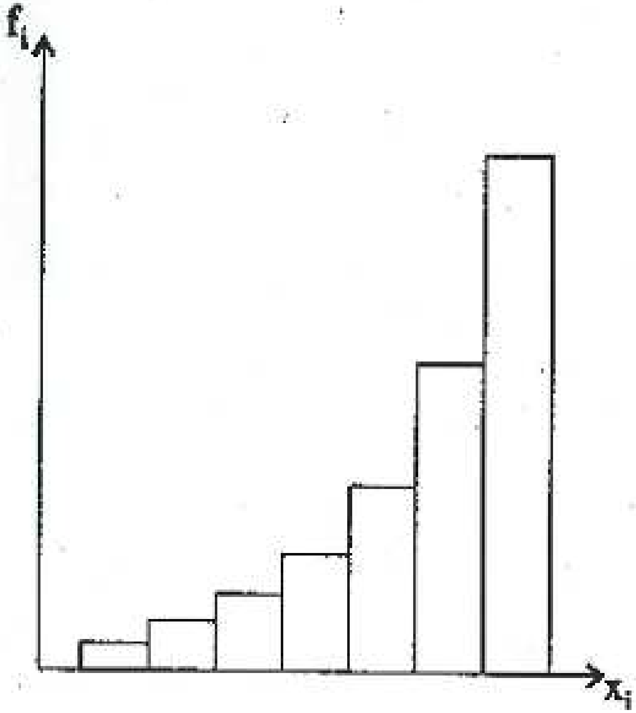
$Mo < Me < \bar{x}$



$\bar{x} < Me < Mo$

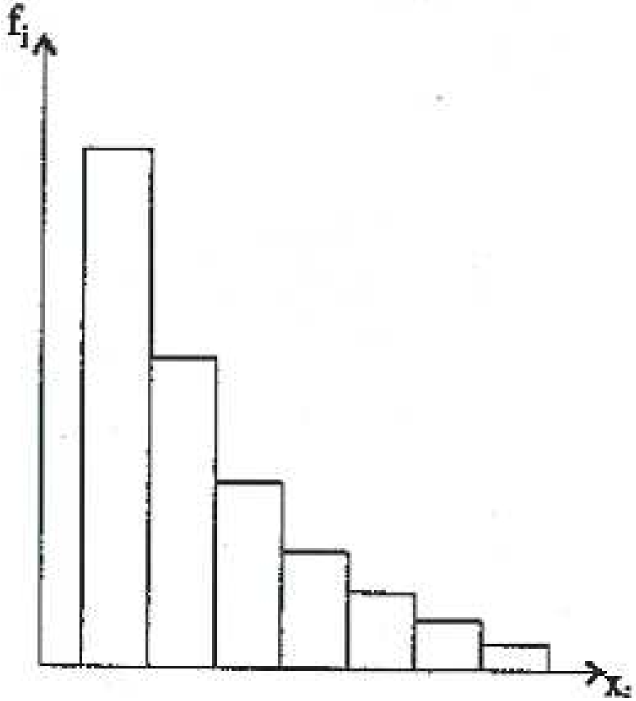
Erősen aszimmetrikus eloszlások

J alakú



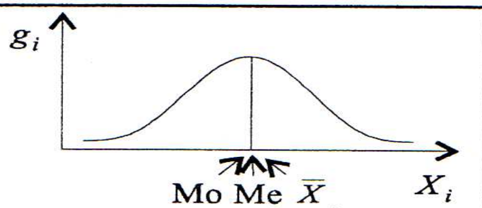
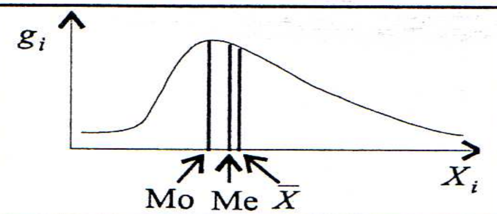
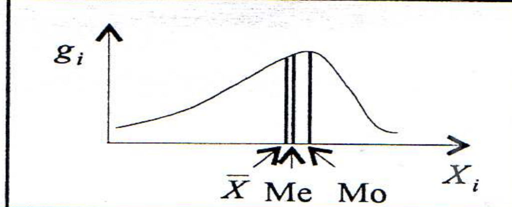
J

Fordított J alakú



ι

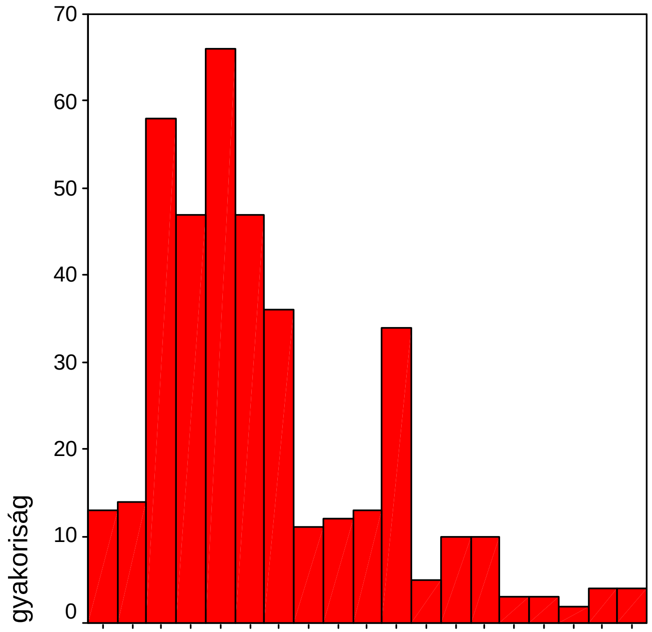
A helyzetmutatók elhelyezkedése szimmetrikus és aszimmetrikus eloszlás esetében

Szimmetrikus eloszlás	Aszimmetrikus eloszlás	
	bal oldali	jobb oldali
		
$Mo = Me = \overline{X}$	$Mo < Me < \overline{X}$	$Mo > Me > \overline{X}$
$(Q_3 - Me) = (Me - Q_1)$	$(Q_3 - Me) > (Me - Q_1)$	$(Q_3 - Me) < (Me - Q_1)$

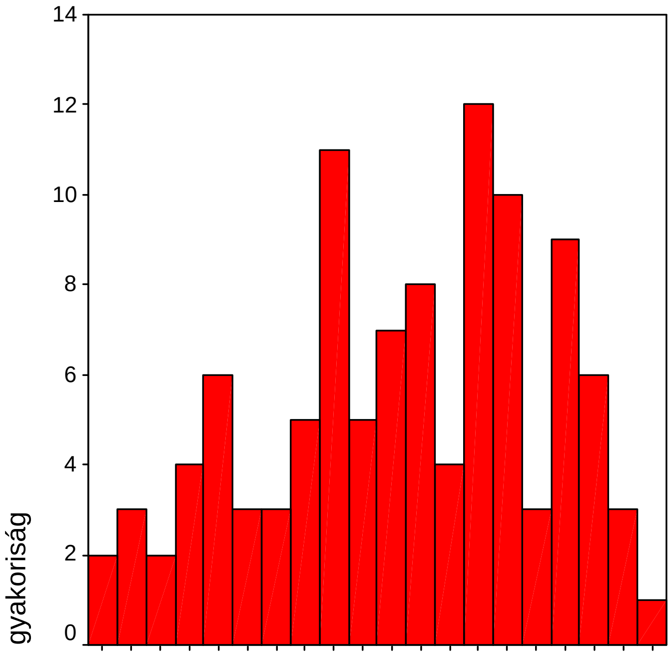
Az eloszlást jellemző statisztikák

$$s = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^3}{\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}\right)^3}$$

FERDESÉG (skewness)



Jobbra ferdül, s > 0

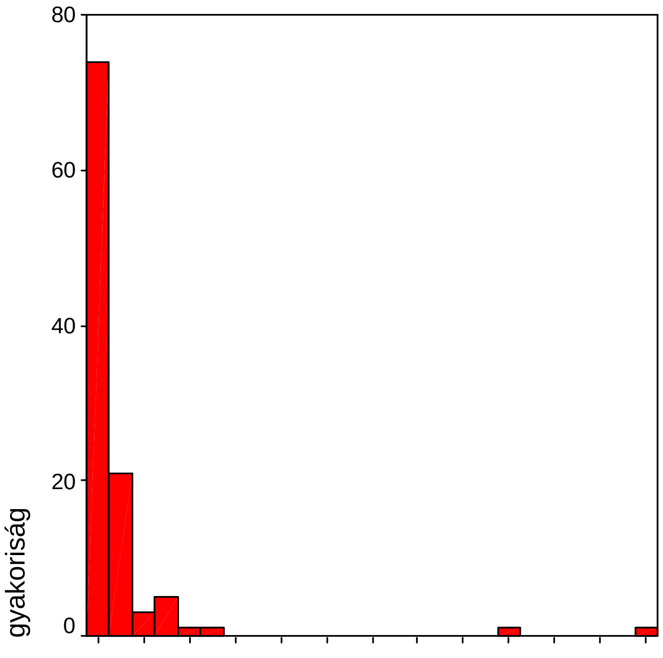


Balra ferdül, s < 0

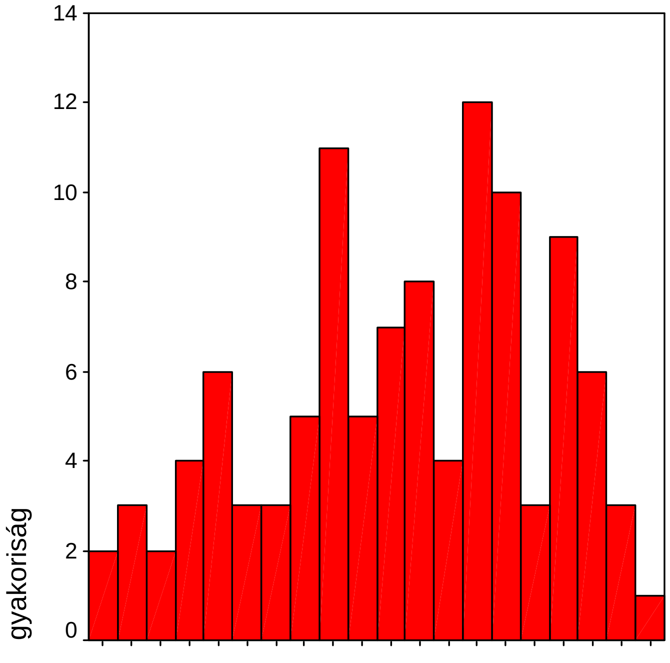
Az eloszlást jellemző statisztikák

LAPULTSÁG (curtosis)

$$k = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^4}{\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}\right)^4} - 3$$



Csúcsos eloszlás, k > 0



Lapos eloszlás, k < 0

Aszimmetria mutatók (1)

Pearson-féle A mutató

Előjele az aszimmetria irányát mutatja.

$A > 0$: bal oldali, jobbra elnyúló aszimmetria

$A < 0$: jobb oldali, balra elnyúló aszimmetria

$A = 0$: szimmetrikus eloszlás.

Abszolút értékének nincs felső korlátja.

- A>1 meglehetősen erős aszimmetria

$$A = \frac{\bar{x} - Mo}{\sigma}$$

Aszimmetria mutatók (2)

Az aszimmetria F-mutatószáma

Kvartiliseken alapul:

$$F = \frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Me) + (Me - Q_1)}$$

Szimmetrikus eloszlás esetén: **F = 0**

Bal oldali aszimmetria esetén: **F > 0**

Jobb oldali aszimmetria esetén: **F < 0**

$-1 \leq F \leq 1$

4) További elemzési módszerek

- Koncentráció
- Idősorok elemzése átlagokkal

Koncentráció

- Gazdasági életben: erőforrások tömörülése, összpontosulása
- Statisztikailag: egy sokaság mennyiség ismérv szerinti vizsgálata
- **Koncentráció:** az értékösszeg jelentős része a sokaság kevés egységére összpontosul

Koncentráció

A koncentráció a relatív gyakoriságok (g_i) és a relatív értékösszegek (z_i) összehasonlításával elemezhető. Ha az egyes osztályközökhöz tartozó g_i és z_i értékek azonosak, az a koncentráció hiányaként értelmezhető, eltérésük viszont a koncentrációt jelzi.

Értékösszeg: $s_i = f_i \cdot x_i$

az adott csoportra jellemző

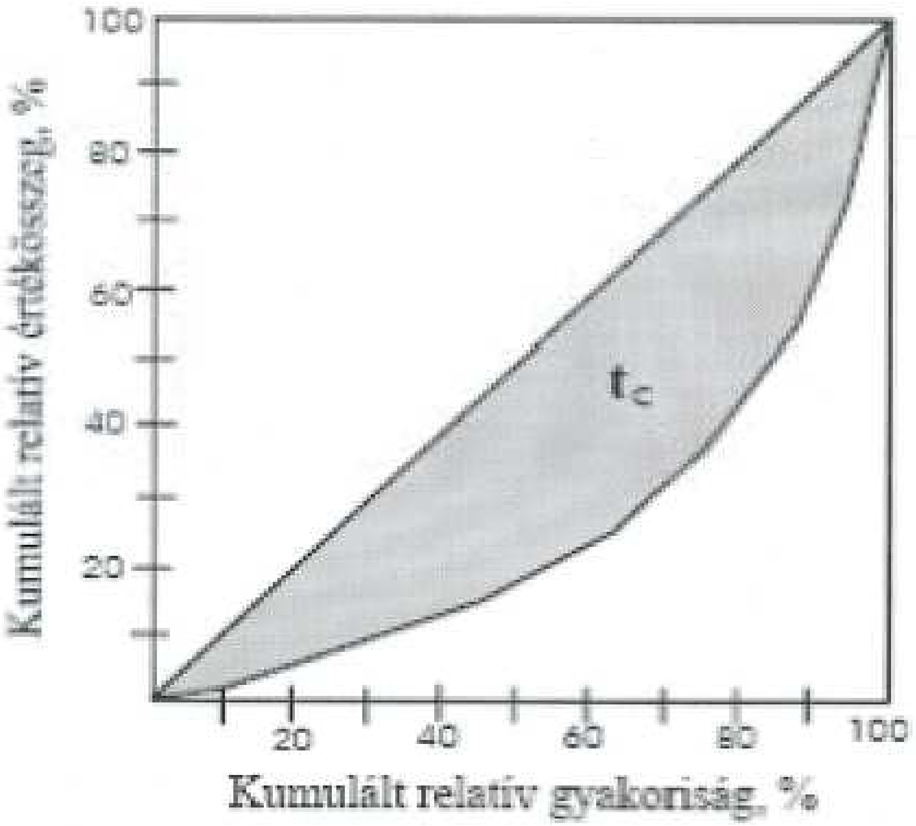
x_i érték (osztályközös gyakorisági sornál az *osztályközép*) és a *gyakoriság* ***f_i*** szorzata

Relatív értékösszeg: $z_i = \dfrac{s_i}{\sum s_i}$

Relatív gyakoriság: $\mathrm{g}_\mathrm{i} = \dfrac{\mathrm{f}_\mathrm{i}}{\sum \mathrm{f}_\mathrm{i}}$

Lorenz-görbe

Egységoldalú négyzetben elhelyezett ábra, amely a kumulált relatív értékösszeg értékeket a kumulált relatív gyakoriságok értékeinek függvényében ábrázolja.



- Ha minden egyed (illetve a sokaság minden részhalmaza egyformán részesedik az értékösszegből, akkor a Lorenz görbe a négyzet átlója lesz; ha egyenlőtlenségek vannak a sokaságban, akkor a Lorenz görbe az átló alatt fut. Minél jobban eltér a görbe lefelé az átlótól, annál nagyobb az adott mennyiség (jövedelem, vagyon, stb.) koncentrációja.

Lorenz-görbe

- Koncentráció hiánya esetén a görbe egybeesik az átlóval.
- Minél távolabb esik a görbe az átlótól, annál nagyobb fokú a koncentráció.

<u>Felhasználása:</u>

- relatív koncentráció szemléltetése
- interpoláció
- több ismérv koncentrációjának egybevetése
- adott ismérv koncentrációjának időbeli vagy térbeli egybevetése

Koncentrációs együttható

A Lorenz-görbe és az átló által bezárt területet koncentrációs területnek nevezzük.

Ha koncentrációs területet a háromszög területéhez viszonyítjuk, akkor e hányados alapján következtetni tudunk a koncentráció mértékére. A koncentrációs terület arányát a koncentrációs együtthatóval mérjük.

$$K = \frac{G}{2\overline{x}}$$

(ahol G: az átlagos különbség (Gini-féle mérőszám))

K értéke [0,1] intervallumban mozoghat, koncentráció hiány esetén K=0, és a K minél közelebb van 1-hez, annál erősebb a koncentráció.

Koncentráció

- Abszolút koncentráció: az értékösszeg kevés egységére összpontosul (pl.: energiaiparban, gépkocsigyártásban)

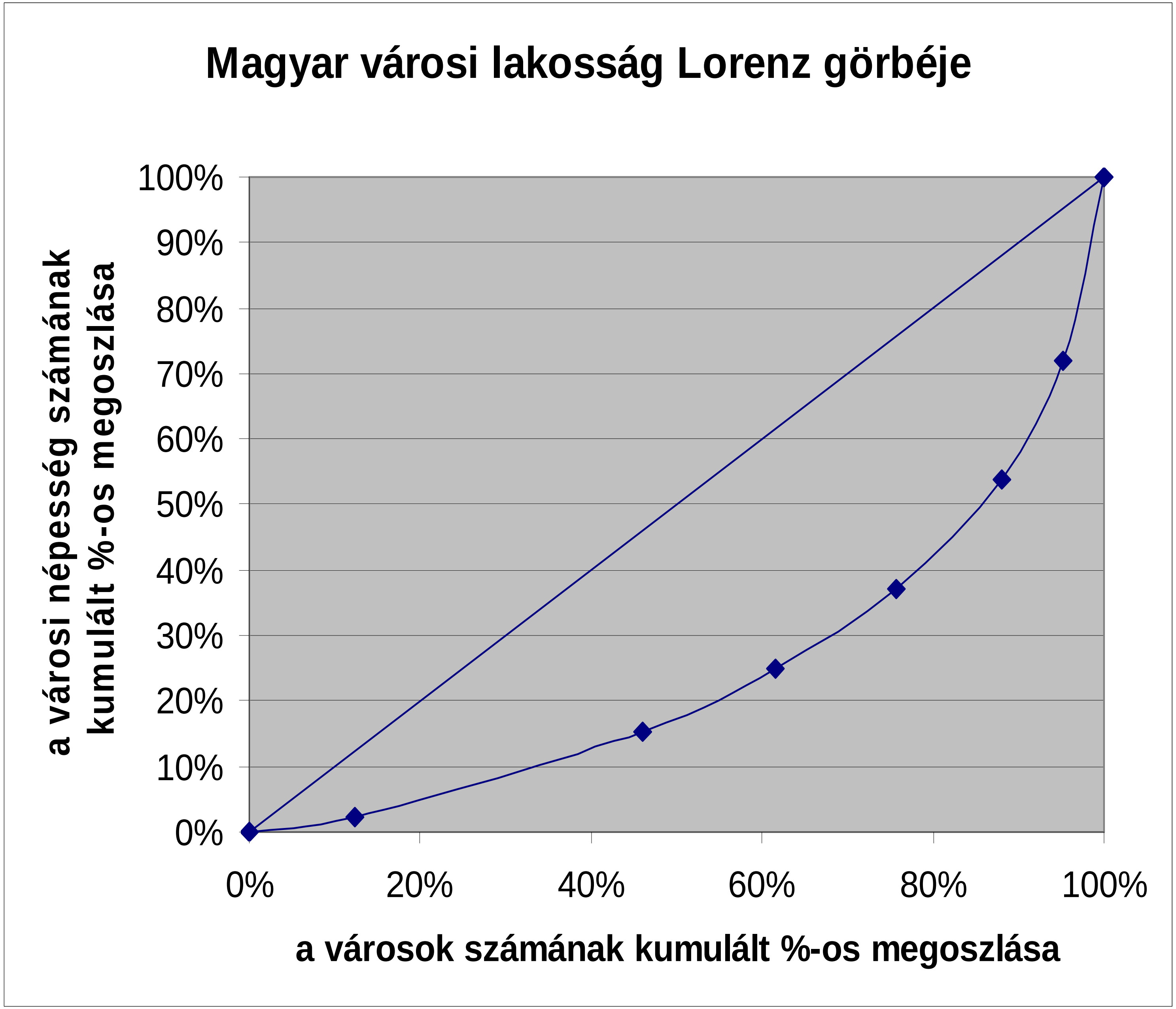
- Relatív koncentráció: az értékösszeg relatív értelemben kevés egységnél összpontosul (pl.: személyi jövedelemben)

Koncentráció

ÉRTÉKÖSSZEG (s)	SOKASÁG (n)
tőke, vagyon, termelés, forgalom, eredmény	gazdasági szervezetek
export, import	országok, termékek, gazdasági szervezetek
mezőgazdasági földterület, eszközállomány, állatállomány	gazdasági szervezetek, tulajdonosok
lakossági jövedelem, vagyon	lakosság, háztartások

Példa:

A népesség száma, fő	Osztályköz hosszúság	A városok			
		száma	számának kumulált megoszlása, %	népessége	népessége kumulált megoszlása, %
- 5999	?	24	12,40	109 216	2,40
6000-11999	6000	65	46,10	580 133	15,30
12000-17999	6000	30	61,60	430 350	24,90
18000-23999	6000	27	75,60	546 159	37,00
24000-39999	16000	24	88,00	751 171	53,70
40000-79999	40000	14	95,30	821 709	72,00
80000-	?	9	100,00	1 261 569	100,00
Összesen :		193		4 500 307	


Magyar városi lakosság Lorenz görbéje
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0%
20%
40%
60%
80%
100%
a városi népesség számának kumulált %-os megoszlása
a városok számának kumulált %-os megoszlása

Példa:

Tapasztalható-e koncentráció a Budapest nélküli városok népességmegoszlásában Magyarországon?

Népességmegoszlás – városok (1997. év végi népességszámok)

Népesség (fő)	**Városok száma**
2 000–4 999	20
5 000–9 999	61
10 000–49 999	105
50 000–99 999	11
100 000–	8
Összesen:	205

Forrás: Magyar Statisztikai Évkönyv 1997.

Számítsuk ki a relatív gyakorisági- és a relatív értékösszeg sor adatait!

relatív gyakoriságok:

Népesség (fő)	**Városok száma (f_i)**	**Relatív gyakoriság, % (g_i)**
2 000–4 999	20	**9,8**
5 000–9 999	61	**29,8**
10 000–49 999	105	**51,1**
50 000–99 999	11	**5,4**
100 000–	8	**3,9**
Összesen:	205	**100,0**

Forrás: Magyar Statisztikai Évkönyv 1997.

A relatív érték-összegek munkatáblája

Népesség (fő)	**Városok száma (f_i)**	**Osztályközép (x_i)**	**Értékösszeg, Lakosság, fő ($f_i x_i$)**	**Relatív értékösszeg, % (z_i)**
2 000–4 999	20	3 500	70 000	**1,3**
5 000–9 999	61	7 500	457 500	**8,3**
10 000–49 999	105	30 000	3 150 000	**57,2**
50 000–99 999	11	75 000	825 000	**15,0**
100 000–	8	125 000	1 000 000	**18,2**
Összesen:	205	–	5 502 500	**100,0**

A kumulált relatív gyakoriságok és értékösszegek

Népesség (fő)	Kumulált relatív	
	Gyakoriság, % (g'_i)	Értékösszeg, % (z'_i)
2 000–4 999	**9,8**	**1,3**
5 000–9 999	**39,6**	**9,6**
10 000–49 999	**90,7**	**66,8**
50 000–99 999	**96,1**	**81,8**
100 000–	**100,0**	**100,0**

A takarékbetét-állomány megoszlása:

Elhelyezett összeg (e Ft)	Betétkönyvek száma (db)
- 100	18
101 - 500	15
501 - 1000	9
1 001 - 3 000	10
3 001 - 5 000	5
5 001 -	3
Összesen:	60

Szemléltesse a betétállomány koncentrációját!

Határozza meg a koncentrációs együtthatót!

Elhelyezett összeg (e Ft)	Betétkönyvek száma (db)	X_i	g_i (%)	S_i	Z_i (%)	g'_i	Z'_i
0 - 100	18	50	30,0	900	1,3	**30,0**	**1,3**
101 - 500	15	300	25,0	4 500	6,4	**55,0**	**7,7**
501 - 1000	9	750	15,0	6 750	9,6	**70,0**	**17,3**
1 001 - 3 000	10	2 000	16,7	20 000	28,5	**86,7**	**45,8**
3 001 - 5 000	5	4 000	8,3	20 000	28,5	**95,0**	**74,3**
5 001 - 7 000	3	6 000	5,0	18 000	25,7	**100,0**	**100,0**
Összesen:	60		100,0	70 150	100,0		

A kumulált relatív gyakorisági értékek jelentősen meghaladják a kumulált relatív értékösszeg adatait $\Rightarrow$ a városi népesség koncentrációja tapasztalható.

- **Erős a koncentráció, ha a sokaság nagy hányadához a teljes értékösszeg kis hányada tartozik, és viszont.**

- Ha az egységeknek az értékösszegből való részesedése egyforma
  $\Rightarrow$ a kumulált relatív gyakoriságok és a kumulált relatív értékösszegek rendre megegyeznek (g'i = z'i). Ez a koncentráció hiányára utal.
  $\Rightarrow$ a görbe a négyzet átlójával egybeesik.

- Teljes koncentráció esetén a görbe egybeesik a koordinátatengelyekkel.

- **A görbe és az átló által bezárt terület a koncentráció relatív nagyságát jellemzi.**

Példa:

Benzin-fogyasztás (l/100 km)	Autók száma
3,0 - 6,0	8
6,1 - 10,0	9
10,1 - 20,0	3
Összesen:	20

Tapasztalható-e koncentráció az egyes kategóriákhoz tartozó autók benzinfogyasztásában?

Nézzük mindig a dolgok napos oldalát!

Mára befejeztük, viszontlátásra!