5. lecke: Aszimmetria, koncentráció

Koncentráció

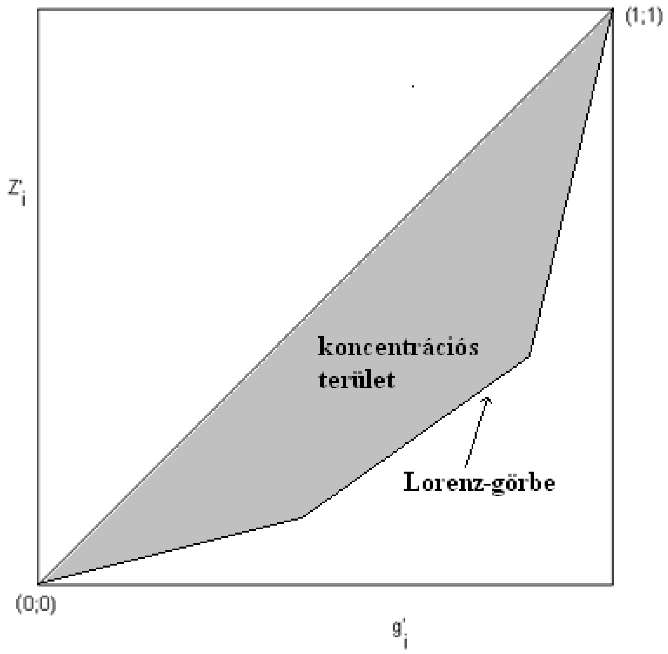
1. Koncentráció vizsgálata

Koncentráció alatt azt értjük, hogy egy változó esetében az értékösszeg nem egyenletesen oszlik meg az egységek (megfigyelések) között, hanem ez néhány egyedre összpontosul. Abban az estben, ha kevés a megfigyelési elemszámunk, akkor az értékösszeg a megfigyelés elemszámából adódóan csak kevés egyedre összpontosulhat. Ezt nevezzük **abszolút koncentrációnak**. Abban az esetben, ha nagy megfigyelési elemszámmal dolgozunk, és ekkor jut az értékösszeg nagy hányada néhány egyedre, akkor **relatív koncentrációról** beszélünk. A koncentráció a szóródás egyfajta megnyilvánulása.

A koncentráció jellemzésére, szemléltetésére számos eljárás ismert, ezek közül csak néhányat említenék meg: kvantiliseloszlás, Lorenz-görbe, Gini együttható (Gini-féle koncentrációs index), Herfindahl-index.

A Lorenz-görbe arra épül, hogy ha nincs koncentráció, akkor $g'_i=Z'_i$. Ha egy vonaldiagramon ábrázoljuk a (0;0) pontot, valamint a felfelé kumulált relatív értékösszegeket a felfelé kumulált relatív gyakoriságok függvényében, akkor megkapjuk a **Lorenz-görbe** pontjait. Tehát a görbe pontjai[1]: (0;0);(g'_i;Z'_i);(1;1). Ha ezt az ábrát egy egységnégyzetben képzeljük el, akkor a koncentráció nem léte azt jelenti, hogy a Lorenz-görbe egybe esik a négyzet átlójával. Minél nagyobb a koncentráció mértéke, a Lorenz-görbe annál jobban eltávolodik a négyzet átlójától. Mivel a mennyiségi sorok ismérvérték szerint növekvő sorrendben használandóak, így a görbe mindig a négyzet átlója alatt marad. A Lorenz-görbe és a négyzet átlója közötti területet koncentrációs területnek nevezzük.

Lorenz-görbe



A Lorenz-görbe nevezetes pontja az **átlagpont**, amelynek vízszintes koordinátája az átlagnál kisebb egységek sokaságon belüli arányát, függőleges koordinátája pedig az átlagnál kisebb egységekhez tartozó értékösszeg hányadát adja meg. Az átlagpont a Lorenz-görbének az a pontja, amelyben egy, a négyzet átlójával párhuzamos egyenes érinti a görbét.

A koncentráció fokát mérhetjük a koncentrációs terület és az átló alatti terület hányadosával:

[1] Természetesen dolgozhatunk százalékos formában is, ekkor az utolsó pont: (100%, 100%).

$$L = \frac{t_c}{\frac{1}{2}} = 2 \cdot t_c \, .$$

A koncentráció fokát egyszerűbben számszerűsíthetjük **a Gini-együttható (Gini-féle koncentrációs index)** segítségével is, melynek értéke megegyezik a koncentrációs terület és az átló alatti terület hányadosával.

$$L = \frac{G}{2 \cdot \bar{x}}$$

A mérőszám értéke 0 és 1 közötti. Minél nagyobb a koncentráció foka, annál nagyobb a mutató értéke. A képlet számlálójában a szóródás mérőszámainál megismert átlagos abszolút különbség szerepel.

A **Herfindahl-index** értékét az

$$HI = \sum_{i=1}^{N} Z_i^2$$

összefüggés alapján számíthatjuk ki. Ha nincs koncentráció, akkor minden egyednek ugyanakkora a részesedése a teljes értékösszegből, ekkor a Herfindahl-index értéke 1/N. A mérőszám értéke[2] 1/N és 1 közötti. Minél nagyobb a koncentráció foka, annál nagyobb a mutató értéke. Ezt a mutatót használják (például OECD, FED) annak eldöntésére, hogy egy adott iparág, szektor koncentrált vagy sem. A mutató ezen alakja miatt sokan azt mondják, hogy a Herfindahl-index elsősorban az abszolút koncentrációt méri. Azonban a vizsgált változó relatív szórásának segítségével is kiszámíthatjuk a Herfindahl-index értékét

$$HI = \sum_{i=1}^{N} Z_i^2 = \frac{\mathrm{v}^2 + 1}{N},$$

emiatt tartják sokan alkalmasnak a relatív koncentráció mérésére.

Mivel a Herfindahl-index értéke $1/N$ és 1 közé esik, így különböző adathalmazok összehasonlítása végett normalizálni szokták. A **normalizált Herfindahl-indexet** a

$$HI^* = \frac{HI - \frac{1}{N}}{1 - \frac{1}{N}}$$

összefüggéssel számíthatjuk ki. Minél nagyobb a mutató értéke, annál nagyobb a koncentráció mértéke. Bizonyos területeken 0,18 feletti értéket erős, 0,1 alatti értéket gyenge koncentrációnak tekintenek.

A kvantilisek alapján lehetőségünk van úgynevezett **kvantiliseloszlás** készítésére. Ez a technika egy olyan osztályközös gyakorisági sor elkészítését jelenti, melyben mindenegyes

[2] A mérőszámot szokták a relatív értékösszegek százalékos alakjából is számítani, bár ez statisztikailag nem teljesen szabályos. Ekkor a mérőszám maximális értéke 10000= (100%)*(100%).

osztályban ugyanannyi elem van, azaz minden osztály relatív gyakorisága $1/k$. Ennek általános szerkezete az alábbi.

Kvantiliseloszlás

Osztályköz	Gyakoriság f_i	Relatív gyakoriság g_i
x_{min} – $x_{1/k}$	N/k	$1/k$
$x_{1/k}$ – $x_{2/k}$	N/k	$1/k$
$x_{2/k}$ – $x_{3/k}$	N/k	$1/k$
. . .	. . .	. .
$x_{(k-1)/k}$ – x_{max}	N/k	$1/k$
Összesen	N	1,00

Speciálisan például a kvartiliseloszlás szerkezete az alábbi.

Kvartiliseloszlás

Osztályköz	Gyakoriság f_i	Relatív gyakoriság g_i
x_{min} – Q_1	$N/4$	¼
Q_1 – Q_2	$N/4$	¼
Q_2 – Q_3	$N/4$	¼
Q_3 – x_{max}	$N/4$	¼
Összesen	N	1,00

A kvantiliseloszlások egy tipikus alkalmazása a jövedelemstatisztikában fordul elő, amikor is a jövedelmek alapján a népességet tizedekre osztják (deciliseloszlás), és ezen tizedek jellemzőit külön-külön is megvizsgálják. Például, a szegénység mérésének egy mutatója az első és a tízedik decilis átlagjövedelmeinek hányadosa.

A koncentráció mérőszámai egy érdekes **tulajdonsággal** rendelkeznek. Ha mindenegyes ismérvértéket ugyanazzal a nullától különböző *A* számmal megszorzunk, akkor mérőszámainak értékei nem változnak.

2. Mintafeladatok

1. feladat

Ismertek egy település lakóira vonatkozó adatok.

A családok megoszlása gyermekszám szerint

Gyerekszám (fő)	Családok száma (család)
0	890
1	950

2	650
3	140
4	40
5	15
6	1
Összesen	2 686

Forrás: fiktív

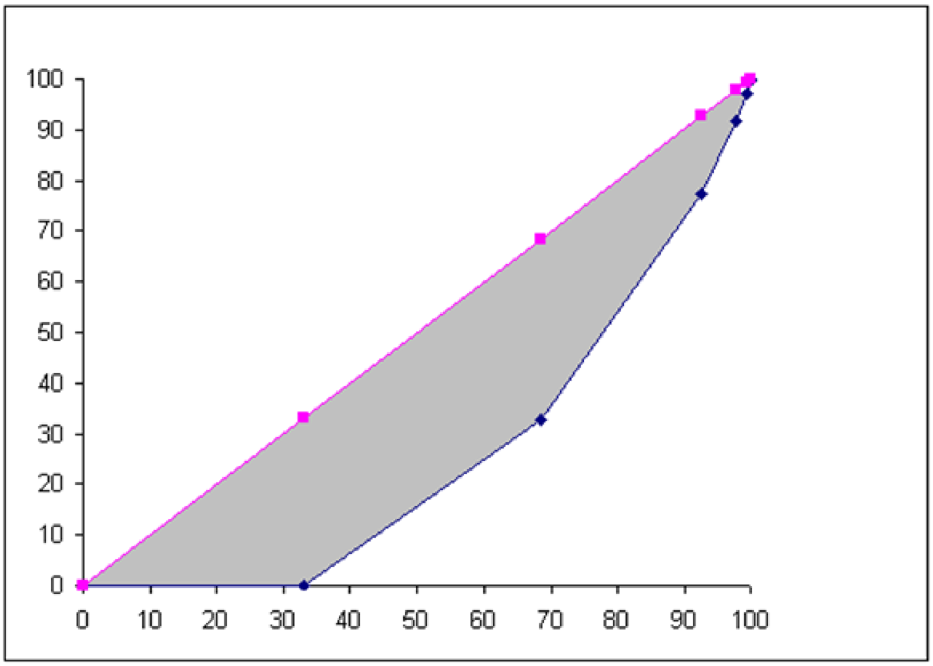
Készítsen Lorenz-görbét! Adja meg a görbe átlagpontjának jelentését! Számítsa ki és értelmezze a Gini együtthatót! (G=1,06)

Lorenz-görbe készítéséhez szükségünk van a felfelé kumulált relatív gyakoriságokra és a felfelé kumulált relatív értékösszegekre.

x_i	f_i	g_i(%)	**g'_i (%)**	S_i	Z_i (%)	**Z'_i (%)**
0	890	33,13	**33,13**	0,00	0,00	**0,00**
1	950	35,37	**68,50**	950,00	32,63	**32,63**
2	650	24,20	**92,70**	1300,00	44,66	**77,29**
3	140	5,21	**97,91**	420,00	14,43	**91,72**
4	40	1,49	**99,40**	160,00	5,50	**97,22**
5	15	0,56	**99,96**	75,00	2,58	**99,80**
6	1	0,04	**100,00**	6,00	0,22	**100,00**
Összesen	2686	100,00	-	2 911	100,00	-

Ez alapján felrajzolhatjuk a Lorenz-görbét.

Lorenz-görbe



A koncentráció jellemzésére vagy a Gini együtthatót használva kapjuk:

$$\mathrm{L} = \frac{G}{2\bar{x}} = \frac{1{,}06}{2 \cdot 1{,}0838} = 0{,}489\,.$$

A mutatók alapján megállapíthatjuk, hogy a sokaság koncentráltsága közepes.

2. feladat

Két megyében külön-külön ismert egy szolgáltatás tekintetében négy vállalat piaci részesedése. Mennyire koncentrált a szolgáltatás piaca?

Vállalat	Piaci részesedés (%)	
	A megye	B megye
1	30	70
2	20	5
3	22	15
4	28	10

Mivel gyakorlatilag a relatív értékösszegek adottak, így alkalmazzuk a Herfindahl-indexet, illetve a normailizált Herfindahl-indexet.

$$HI_A = \sum_{i=1}^{N} Z_i^2 = 0{,}3^2 + 0{,}2^2 + 0{,}22^2 + 0{,}28^2 = 0{,}2568$$

$$HI_A^* = \frac{0{,}2568 - \frac{1}{4}}{1 - \frac{1}{4}} = 0{,}009$$

$$HI_B = \sum_{i=1}^{N} Z_i^2 = 0{,}7^2 + 0{,}05^2 + 0{,}15^2 + 0{,}1^2 = 0{,}525$$

$$HI_B^* = \frac{0{,}525 - \frac{1}{4}}{1 - \frac{1}{4}} = 0{,}37$$

Ez alapján megállapítható, hogy A megyében a szolgáltatás piaca nem koncentrált, míg B megyében igen.

3. Ellenőrző kérdések

1. Mi a koncentráció?
2. Mi a relatív koncentráció?
3. Mi az abszolút koncentráció?
4. Hogyan lehet a koncentrációt vizsgálni?
5. Hogyan lehet a Lorenz-görbét elkészíteni? Mire használható a görbe?
6. Igaz-e, hogy a koncentráció nagysága a Lorenz-görbe alatti területtel jellemezhető?
7. Mi a koncentrációs terület?
8. Mi a Lorenz-görbe átlagpontja?
9. Mit fejez ki a Gini együttható?
10. Mi a Herfindahl-index? Mire használható?
11. Mi a normált Herfindahl-index?
12. Milyen tulajdonságokkal rendelkeznek a koncentráció taglalt mérőszámai?

Aszimmetria

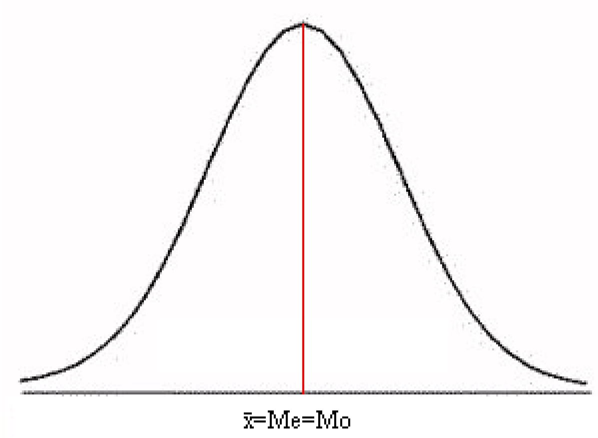
1. Az aszimmetria vizsgálata

Miután a mennyiségi változónkból kiszámítottuk a középértékeket és a szóródást, tudjuk, hogy mennyi a változó átlagos értéke, illetve, hogy az egyes ismérvértékek átlagosan mennyire térnek el az átlagostól, azonban arról még nincs információnk, hogy az értékek többsége átlag alatti vagy feletti, vagy esetleg az ismérvértékek szimmetrikusan helyezkednek el az átlag körül. Például az emberek többségének a havi bruttó keresete a havi bruttó átlagkereset alatti vagy feletti? Ezt a kérdéskört vizsgálják az **aszimmetria (**angolul skewness**)** mérőszámai.

Ha egy eloszlásnak több módusza van, akkor azt **többmóduszú** eloszlásnak nevezzük. Például a kétmóduszú eloszlások tipikus grafikus képe egy *M, U* vagy *W* betűre emlékeztet. A továbbiakban az egymóduszú eloszlásokkal foglalkozunk.

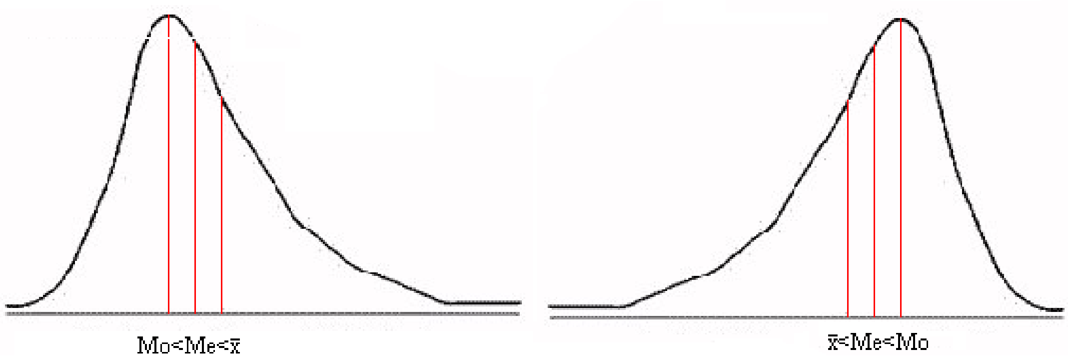
Ha az eloszlás szimmetrikus, akkor a módusz, a medián és a számtani átlag értéke azonos. A szimmetrikus eloszlás tipikus ábrája az alábbi.

Szimmetrikus eloszlású változó



Ha az eloszlás aszimmetrikus, akkor a módusz, a medián és a számtani átlag értéke nem azonos. Ekkor két esetet különböztethetünk meg: a módusz és a medián is kisebb az átlagnál, vagy mindkettő nagyobb[3]. **Ha a módusz és a medián is kisebb az átlagnál, az azt jelenti, hogy az ismérvértékek többsége kisebb az átlagnál.** Ezt az esetet **baloldali aszimmetriának** nevezzük. **Ha a módusz és a medián is nagyobb az átlagnál, az azt jelenti, hogy az ismérvértékek többsége nagyobb az átlagnál.** Ezt az esetet **jobboldali aszimmetriának** nevezzük. A bal és jobboldali aszimmetria elnevezés a szakirodalomban nem egységes, van, aki fordítva nevezi el a két típust, attól függően, hogy a módusz vagy pedig az eloszlás hosszan elnyíló részére jellemző elmozdulási irányt kívánják jellemezni. A lényeg nem az elnevezésen van, hanem azon, hogy el tudjuk dönteni azt, hogy az ismérvértékek többsége átlag alatti vagy feletti.

Aszimmetrikus eloszlású változó



Az aszimmetria megállapításához vagy grafikus ábrát készítünk, vagy valamilyen mérőszámot alkalmazunk. A grafikus ábrák esetében csak akkor vonhatunk le megfelelő következtetést, ha az ábrán a módusz, medián és számtani átlag értéke is jelölve van.

[3] Empirikus tapasztalatok azt mutatják, hogy nagy elemszámú megfigyelések során, aszimmetria esetében $Mo < Me < \bar{x}$, vagy $Mo > Me > \bar{x}$.

Az aszimmetria mérőszámainak[4] közös jellemzője, hogy negatív értékük arra utal, hogy az ismérvértékek többsége magasabb az átlagosnál, míg pozitív értékük azt jelzi, hogy az ismérvértékek többsége alacsonyabb az átlagosnál.

Az aszimmetria egy mérőszáma a medián és számtani átlag közötti relációra épül:

$$P = 3 \cdot \frac{\bar{x} - Me}{\sigma}.$$

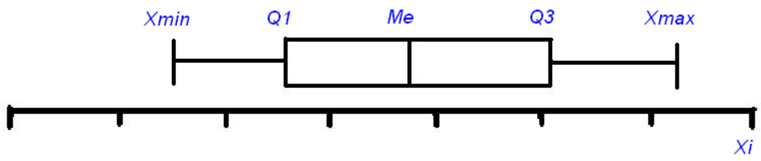
Az aszimmetria egy másik mérőszáma a kvartilisekre épít. Ugyanis, ha egy eloszlás szimmetrikus, akkor a medián (második kvartilis) egyenlő távolságra helyezkedik el az első és a harmadik kvartilistől $(Q_3 - Me) = (Me - Q_1)$. Ezekre a különbségekre építve felírható az

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

képlet. Az F mutató értéke -1 és +1 közé esik. Minél messzebb van a mutató értéke a nullától, annál erősebb az aszimmetria mértéke.

A kvartilisek alapján elkészíthető az úgynevezett **boxplot**. Ez egy olyan grafikus ábrát jelent, melyen a legkisebb és a legnagyobb ismérvérték egy-egy talppal, míg a kvartilisek egy-egy dobozfallal vannak jelölve. Az ábra alapján sejtéseket fogalmazhatunk meg az aszimmetriára vonatkozóan abból, hogy a medián a doboz mely végéhez (Q_1 vagy Q_3) van közelebb, ugyanis ezek a távolságok szerepelnek az F mutatóban. Ha a medián a doboz Q_1 végéhez van közelebb, akkor az ismérvértékek többsége kisebb az átlagosnál.

Boxplot



Az aszimmetria és a csúcsosság mérőszámai egy érdekes **tulajdonsággal** rendelkeznek. Ha mindenegyes ismérvértéket ugyanazzal az A számmal növelünk vagy egy nullától különböző A számmal megszorzunk, az aszimmetria mérőszámainak értékei nem változnak.

[4] A számítógépes szoftverek egy más típusú mérőszámot alkalmaznak a tárgyaltakhoz képest. Ennek alapja az $\alpha_3 = \frac{\sum_{i=1}^{N} d_i^3}{N \cdot \sigma^3}$ mérőszám.

2. Mintafeladat

Jellemezze egy vizsgált sokaság eloszlását az alábbi adatok ismeretében!

$D_1 = 3{,}22$
$D_9 = 15{,}6$
$Q_1 = 6{,}62$
$Q_3 = 10{,}2$
$Q_2 = 7{,}4$

Vizsgáljuk meg a sokaság eloszlásának aszimmetriáját!

$$F = \frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Me) + (Me - Q_1)} = \frac{(10{,}2 - 7{,}4) - (7{,}4 - 6{,}62)}{(10{,}2 - 7{,}4) + (7{,}4 - 6{,}62)} = 0{,}56$$

Ezek szerint a sokaság eloszlása – az azonos szórású normális eloszlás gyakorisági görbéjéhez képest –baloldali aszimmetriát mutat, azaz az értékek többsége átlag alatti.

3. Ellenőrző kérdések

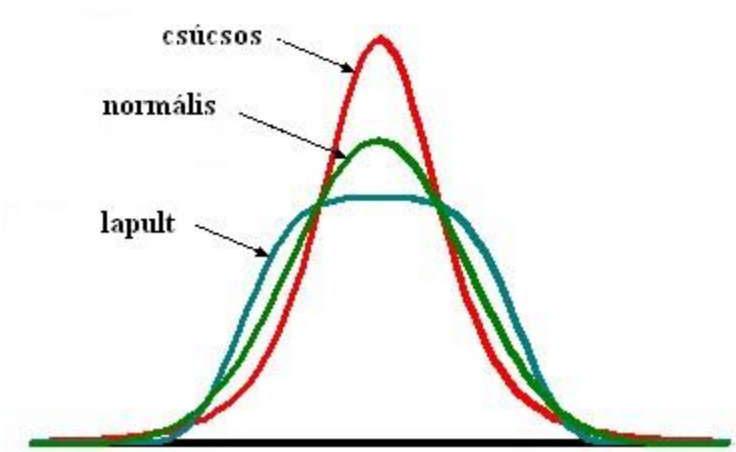
1. Mire ad választ az aszimmetria vizsgálata?
2. Grafikus ábrázolás alapján milyen sejtéseket lehet megfogalmazni egy egymóduszú sokaság aszimmetriájára vonatkozóan?
3. Hogyan értelmezhetőek az aszimmetria mérőszámai?
4. Hogyan értelmezné a P=0,6 értéket?
5. Hogyan értelmezné a F=-0,1 értéket?
6. Hogyan értelmezné az α_3=0,1 értéket?
7. Milyen tulajdonsággal rendelkeznek az aszimmetria mérőszámai?

Kiegészítő olvasnivaló (nem képezi részét a számonkérésnek!)

1. A csúcsosság vizsgálata

Az aszimmetria mellett vizsgálhatjuk az eloszlások **csúcsosságát**, illetve lapultságát. Ennek esetei az alábbi ábrán láthatóak.

Eloszlások csúcsossága



Forrás: http://grants.hhp.coe.uh.edu/doconnor/PEP6305/KurtosisPict.jpg

A csúcsosságot, az aszimmetriához hasonlóan mindig valamihez viszonyítva adhatjuk meg. A viszonyítás alapja mindig az azonos szórású normális eloszlás gyakorisági görbéje. A csúcsosság egyik mérőszáma a K-mutató, mely arra épül, hogy csúcsos eloszlások esetében az interkvartilis terjedelem az interdecilis terjedelemhez képest kisebb a lapultabb eloszlásokhoz viszonyítva.

$$K = \frac{Q_3 - Q_1}{2(D_9 - D_1)}$$

A normális eloszlás esetében a mutató értéke 0,263. Az azonos szórású normális eloszláshoz képest lapultabb eloszlások esetén K>0,263, csúcsosabb eloszlások esetében K<0,263.

A csúcsosság egy másik lehetséges mérőszáma a szoftverekben is előforduló

$$\alpha_4 = \frac{\sum_{i=1}^{N} d_i^4}{N \cdot \sigma^4} - 3.$$

Az azonos szórású normális eloszláshoz képest lapultabb eloszlások esetén a mutató értéke negatív, csúcsosabb eloszlások esetében a mutató értéke pozitív.

Az aszimmetria és csúcsosság jellemzőinek együttes figyelembevételével érdekes következtetéseket vonhatunk le. Például egy szimmetrikus, csúcsos eloszlás azt jelenti, hogy az átlaghoz közeli értékek fordulnak elő nagy valószínűséggel, azaz az átlag körül tömörülnek az ismérvértékek. Egy szimmetrikus, lapult eloszlás azt jelenti, hogy az ismérvértékek közel egyenletesen helyezkednek el az átlag körül.

A csúcsosság mérőszámai –az aszimmetriához hasonlóan- egy érdekes **tulajdonsággal** rendelkeznek. Ha mindenegyes ismérvértéket ugyanazzal az *A* számmal növelünk vagy egy nullától különböző *A* számmal megszorzunk, a csúcsosság mérőszámainak értékei nem változnak.