

A negyedik mutatóról

Hunyadi László

egyetemi tanár

E-mail: lhunyadi@chello.hu

A leíró statisztika az egydimenziós eloszlások jellemzésére általában négy mutatószámot használ: a helyzetet, a szóródást, a szimmetriát/aszimmetriát, valamint a normális eloszláshoz viszonyított alakot jellemző mutatókat. Ez a tanulmány a negyedik típussal, a lapultságot, illetve csúcsosságot kifejező alakmutatókkal foglalkozik.

Fő mondanivalója az, hogy maga a fogalom (kurtózis) meglehetősen ellentmondásos, ezért nem célszerű egyetlen mutatószámmal jellemezni. A cikk egy sor lehetséges mérőszámot definiál, majd azok tulajdonságait néhány, a társadalmi és gazdasági elemzésekben gyakran használt nevezetes eloszlást alapul véve összehasonlító módon elemzi. Rámutat arra, hogy ezek a mérőszámok gyakran elég gyengén korrelálnak egymással, így a magas mérési skálán mért mutatók helyett olykor elegendő az eloszlások kurtózis szerinti rangsorolása. A tanulmány arra is felhívja a figyelmet, hogy ezen mutatók jó része szoros kapcsolatban áll a normalitás mintából való ellenőrzésére szolgáló statisztikai próbákkal.

TÁRGYSZÓ:

Statisztikai módszertan.

Valószínűség-eloszlások.

A gyakorisági eloszlások jellemzésekor a statisztikában elterjedt szokássá vált, hogy az eloszlások közvetítette információkat lényegében négy mutatóban próbálják meg összesűriteni. Ezek közül első valamelyik középérték (amit egyesek a central tendency közvetlen fordításaként centrális tendenciaként említenek), a második a különbözőséget jellemző valamely szóródás mutató (leggyakrabban a szórás vagy a variancia), a harmadik az eloszlás szimmetriáját vagy aszimmetriáját jellemző valamely ferdeségi mutató, végül a negyedik mutató a normális eloszláshoz viszonyított alak valamilyen jellemzője. Ezt a negyedik tulajdonságot szokták kurtózisnak nevezni; a magyar szaknyelvben *csúcsosság* vagy *lapultság* néven vált ismertté. Az empirikus elemzések során, főként az említett két első mutatót (középérték és szóródás) használják, jóllehet az eloszlás alakját jellemző harmadik és negyedik mutató fontosságát egy sor gyakorlati probléma (például a jövedelemeloszlások összehasonlítása) is hangsúlyozza.

A kurtózismutató, az ún. *negyedik mutató* vizsgálatunk tárgya. Jóllehet a leíró statisztikai mutatók kutatása nem tartozik a statisztikai kutatások élvonalába, a negyedik mutató vizsgálata nem tűnik érdektelennek. Egyrészt azért mert maga a fogalom (kurtózis) meglehetősen laza, sok értelmezése lehet, nem pontos, másrészt azért, mert éppen ennek következtében a negyedik mutatóval kapcsolatban félreértések fogalmazódnak meg. A tanulmány célja, hogy körüljárja a kurtózis fogalmát, lehetséges mérési módjait, a lehetőségek keretein belül összevesse mutatószámait, azok tulajdonságait, és útmutatót adjon e fogalom és mutatóinak helyes használatához.

Ennek megfelelően a tanulmány felépítése a következő. Először áttekintjük a kurtózis fogalmát és néhány definícióját, majd bemutatjuk néhány ismert, illetve eddig még nem használt mérőszámát. Ezt követően megkíséreljük összevetni a különböző mérőszámokat, illetve bemutatunk egy eljárást, amelyik az egyes eloszlások kurtózis szerinti sorba rendezésére alkalmas. A befejező fejezet a fontosabb tanulságok összegzését, az itt nem tárgyalt fontosabb kérdéseket, valamint a gyakorlati alkalmazás lehetőségeit foglalja össze.

1. A kurtózis fogalma

A kurtózis (kurtosis) szó, melyet először a fogalom megalkotója, *Pearson* használt, a görög *κυρτός*, *kyrtos* vagy *kurtos*, szóból ered, ami (az angol *bulging* szó közvetítésével) kicsúcsosodást, kidülledést, kihasasodást jelent. Ez a fogalom meglehetősen szemléletes, ugyanakkor, mint látni fogjuk némiképp egyoldalú, ezért az elne-

vezésből adódóan is félreértések forrása lehet. Másutt is előforduló probléma ugyanis az, hogy egy elnevezés – éppen köznapi szemléletessége folytán – a fogalom olyan leegyszerűsítésével jár, amely esetenként éppen a lényegét hagyja elsikkadni.

A csúcosság vagy ritkábban lapultság elnevezés, mint a későbbiekben látni fogjuk, nem igazán ragadja meg a jelenség lényegét, ezért egy végső következtetést megelőlegezve azt javasoljuk, hogy a jelenségre használjuk inkább a kissé obskúrus kurtózis szót, és a csúcosságot vagy lapultságot, illetve az ezekkel járó további tulajdonságot csak a magyarázatok során alkalmazzuk. Felfogásunk szerint ugyanis ezek a fogalmak a kurtózisnak csak egy-egy oldalát, és talán nem is leglényegesebbet fejezik ki.

Az etimológián túllépve próbáljuk meg definiálni a kurtózis fogalmát. A következőkben néhány, a szakmában mértékadónak számító forrásból idézzük a kurtózis fogalmát:

„Given two frequency distributions which have the same variability as measured by the standard deviation, they may be relatively more or less flat-topped than the normal curve. If more flat-topped I term them platykurtic, if less flat-topped leptokurtic, and if equally flat-topped mesokurtic” (Pearson [1905]).¹

„Pearson defines kurtosis...as a measure of departure from normality...” (Seier [2055]).²

„...A vizsgált gyakorisági eloszlás ábrájának csúcsa hegyesebb vagy lapultabb lehet, mint a normális eloszlás gyakorisági görbéjének csúcsa... [Ezt az] eltérést csúcosságnak vagy lapultságnak nevezük...” (Hunyadi-Vita [2008]).

„The fourth central moment reflects the peakedness or kurtosis of a probability distribution” (Canavos [1984]).³

„Az X ismérv eloszlásának ... további jellegzetessége, hogy milyen az eloszlás csúcossága (kurtózisa) azaz a módusz környezetében milyen a görbe relatív emelkedése...” (Kerégyártóné et al. [2008]).

„Kurtosis is a measure of a type of departure from normality ... The population kurtosis is given by the fourth moment over the square of the variance...” (Kotz–Johnson [1985]).⁴

¹ „Adott két gyakorisági eloszlás, melyek változékonysága a szórással mérve azonos, ezek viszonylag jobban vagy kevésbé lapult csúcsúak lehetnek, mint a normális görbe. Ha lapultabbak platikurtikusnak, ha kevésbé lapultak leptokurtikusnak, ha azonosan lapultak mezokurtikusnak nevezem őket.”

² „Pearson a kurtózist a normalitástól való eltérés mértékéeként definiálja.”

³ „A negyedik centrális momentum egy valószínűség-eloszlás csúcosságát vagy kurtózisát tükrözi.”

⁴ „A kurtózis a normalitástól való eltérés egy fajtájának mutatója. A sokasági kurtózis a negyedik momentum és a variancia négyzetének hányadosa.”

„Csúcsosságon az eloszlást jellemző görbe meredekségét értjük a módusz környezetében” (Rappai–Pintér [2007]).

„Az, hogy egy statisztikai adathalmazban milyen mértékű az ilyen extrém adatok jelenléte, legjobban az ún. csúcsossági együttható segítségével mérhető...” (Vargha [2000]).

„Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution.” (Pham [2006]).⁵

„In probability theory and statistics, kurtosis ... is a measure of the ‘peakedness’ of the probability distribution of a real-valued random variable” (www.wikipedia.org).⁶

Az idézett források mellett, hogy természetesen valamennyien ugyanarra a jelenségre adnak definíciót vagy éppen magyarázatot, értelmezést, a jelenség más és más oldalát hangsúlyozzák. *Pearson* a szórással mért azonos szóródást, *Seier* a normalitástól való eltérést, *Hunyadi* és *Vita* a görbe csúcsának hegyességét, a Wikipedia lexikon a csúcsosságot, *Canavos* a mérésére szolgáló negyedik centrális momentumot, *Kerékgyártóné és szerzőtársai*, valamint *Rappai* a gyakorisági görbe meredekségét, *Vargha* pedig az extrém adatok előfordulásának gyakoriságát emeli ki. Ez a sokszínűség jól mutatja azt, hogy egy meglehetősen összetett jelenséggel állunk szemben.

Ha ezek után definiálni szeretnénk a kurtózist, akkor a következő tulajdonságokat kellene figyelembe vennünk: a kurtózis egy empirikus vagy elméleti eloszlás tulajdonsága, amelyik

- az azonos szóródású normális eloszlás alakjától való eltérést jellemzi;
- kiemelve az eloszlásnak a módusz körüli kicsúcsosodást vagy el-laposodását, valamint
- a szárnyak (tails) vastagságát és hosszát.

Ez a korántsem kizárólagos definíció jól mutatja a fogalom ellentmondásosságát: azt, hogy valójában egy olyan fogalmat szeretnénk megragadni, mérni, különböző eloszlásokon összehasonlítani, ami ily módon nem mérhető. Legalábbis egy vagy néhány mutatószámmal nem jól jellemezhető. Gondoljunk arra az analógiára, hogy egy emberi arcot kellene egy vagy néhány mutatóval leírni! Nyilván lehet különböző

⁵ „A kurtózis azt méri, hogy az adatok eloszlása a normális eloszláshoz viszonyítva csúcsosabb, vagy lapultabb.”

⁶ „A valószínűség-elméletben és a statisztikában a kurtózis ... egy valószínűség-eloszlásának „csúcsosságát” méri.”

arányokat, méreteket vizsgálni, de ahhoz, hogy viszonylag pontos leírást tudjunk adni, nagyon sok mérőszám összességére és rendszerére van szükségünk. Egy eloszlás alakjának mérése ennél talán egyszerűbb feladat, de itt is az a probléma, hogy valójában egy minőségi jellemzőt (alak) kell mennyiségi jellemzőkkel közelíteni.

A statisztikusnak azonban az a feladata, hogy számokkal jellemezze a jelenségeket, így nem lehet meghátrálni a kurtózis mérésének problémáitól. A méréssel kapcsolatos kérdések legfontosabbika talán az, hogy mérhető-e a kurtózis magas szintű (intervallum- vagy arányskálán) skálán, vagy csak ordinális szintű mérése célszerű. Erre a kérdésre nyilván csak a megfelelő számítások birtokában lehet válaszolni, ezért a továbbiakban azt az utat követjük, hogy először megpróbáljuk áttekinteni azokat a legfontosabb mutatószámokat, amelyek segítségével a kurtózis jellemezhető, majd ezt követően ezek összevetéséből kísérünk meg következtetéseket levonni.

2. A kurtózis mérőszámairól

Mielőtt megkíséreljük sorra venni a kurtózis lehetséges mérőszámait, egy kicsit pontosítani és szűkíteni kell a vizsgálat tárgyát.

Az első kérdés, ami ennek kapcsán felmerül, hogy az empirikus vagy az elméleti eloszlások mérését tekintjük elsődlegesnek. Kétségtelen, hogy a statisztikai gyakorlatban az empirikus eloszlások alakjának jellemzése az elsődleges feladat, ám empirikus eloszlások mutatószámainak összevetése parttalan feladatnak látszik, ezért a továbbiakban elsősorban elméleti valószínűség-eloszlások esetére származtatjuk, értelmezzük és hasonlítjuk össze a különféle kurtózis mutatókat.

További kérdések merülhetnek fel az eloszlások megválasztásakor. Ezek egyike az, hogy diszkrét vagy folytonos eloszlásokat vizsgálunk, a másik az, hogy csak a szimmetrikus eloszlásokra koncentrálunk, avagy megengedjük-e a nemszimmetrikus eloszlásokat is. Ami a diszkrét-folytonos problémát illeti, ez nem annyira elvi, mint inkább gyakorlati kérdés, hiszen a diszkrét eloszlások esetén az egyes mutatók számításához szükséges kvantilisek értékei gyakran csak nagyon pontatlanul határozhatók meg, ezért a belőlük számított mutatók is pontatlanok lesznek. A szimmetrikus/nemszimmetrikus eloszlások kérdése azért merülhet fel, mert ha a kurtózist, mint a normális eloszlástól vett eltérésként értelmezzük, akkor valójában nincs sok értelme nemszimmetrikus eloszlásokat vizsgálni. Ugyanakkor, ha a csúcsosságot vagy éppen a széles szárnyakat emeljük ki a definícióból, akkor a nemszimmetrikus eloszlások esete éppen olyan fontos, mint a szimmetrikusoké. Ha pedig arra gondolunk, hogy a végső cél mégiscsak empirikus eloszlások jellemzé-

se, akkor ismét megengedhetjük a nemszimmetrikus eloszlásokat, hiszen egy empirikus eloszlás esetén maga a szimmetria egyáltalán nem triviális, jól látható, meg tapasztalható tulajdonság.

Leszűkítjük a vizsgálatunkat abból a szempontból is, hogy nem kívánunk foglalkozni azzal a kérdéssel, hogy miként becsülhetők az egyes elméleti eloszlások kurtózis mutatói mintából, illetve ezeknek a becsléseknek milyen tulajdonságaik vannak. Ez az egyébként igen érdekes kérdés nagyon messzire vezet, ezért most átmenetileg kizártuk a vizsgálódás köréből. Végül az elemzést leegyszerűsítjük annyiban, hogy nem foglalkozunk a kurtózis többdimenziós általánosításaival, hanem pusztán egydimenziós eloszlások esetét vizsgáljuk.

Az eloszlások kiválasztásakor a döntő szempont az volt, hogy olyan eloszlásokat vizsgáljunk, amelyek a társadalmi-gazdasági jelenségeket vizsgáló statisztikai gyakorlatban a leginkább előfordulnak, viszonylag ismertek és kezelésük nem vezet különleges bonyodalmakhoz. Ennek megfelelően az 1. táblázatban megjelenő eloszlások kurtózisainak mutatószámait elemeztük.

1. táblázat

A vizsgált eloszlások csoportosítása

Az eloszlás szimmetriája	Folytonos	Diszkrét
	eloszlások	
Szimmetrikus eloszlások	U, TR, N, LOG, ST	BI($P = 0,5$)
Nemszimmetrikus eloszlások	LN, EXP, χ^2	BI($P \neq 0,5$), PO

Megjegyzés. A rövidítések jelentése: U = egyenletes, TR = háromszög, N = normális, LOG = logisztikus, ST = Student (t), LN = lognormális, EXP = exponenciális, χ^2 = khi-négyzet, BI = binomiális, PO = Poisson.

Az 1. táblázatban megjelenő eloszlások paraméterezése szintén fontos kérdés. Vannak ugyanis – mint látni fogjuk – olyan mutatók, amelyek bizonyos eloszlásokon paraméterfüggetlenek, de a mutatók (és az eloszlások) nagyobb része nem ilyen. Ezért már az eloszlások paraméterezésekor arra törekedtünk, hogy eleve olyan eloszlásokat vizsgáljunk és hasonlítsunk össze, amelyek legfontosabb jellemzői (első és második centrális momentumai) megegyeznek. Ezért valamennyi vizsgált eloszlásnál rögzítettük a variancia értékét ($Var(Y)=2$), és ahol lehetett⁷ a várható értéket is ($E(Y)=0$). Ilyen peremfeltételek mellett az egyes eloszlások paraméterezését a 2. táblázat mutatja.

⁷ Ismeretes, hogy bizonyos eloszlásoknál a várható érték és a variancia nem függetlenek egymástól. Ezek esetében (például exponenciális eloszlás) ez nem tehető meg.

2. táblázat

A vizsgált eloszlások paraméterezése

Eloszlás	Paraméter/szabadságfok	Várható érték	Variancia
Egyenletes	$\alpha = -2,45, \beta = 2,45$	0	2,00
Háromszög	$\alpha = -3,46, \beta = 3,46$	0	2,00
Normális	$\mu = 0, \sigma^2 = 2$	0	2
Logisztikus	$\mu = 0, s^2 = \frac{6}{\pi^2}$	0	2
Student	$\nu = 4$	0	2
Lognormális	$\mu = 0, \sigma^2 = \ln 2$	$\sqrt{2}$	2
Exponenciális	$\theta = \sqrt{2}$	$\sqrt{2}$	2
Khi-négyzet	$\nu = 1$	1	2
Binomiális	$P = 0,1, n = 22$	2,2	1,98
	$P = 0,5, n = 8$	4	2
	$P = 0,8, n = 13$	10,4	2,08
Poisson	$\lambda = 2$	2	2

A kurtózis mérőszámait különféle elvek alapján származtathatjuk, és a továbbiakban ezek szerinti csoportosításban tárgyaljuk. A kurtózist mérhetjük tehát

- momentumok útján;
- kvantilisekkel;
- a szórás és az átlagos abszolút eltérés összevetésével;
- információelméleti mérőszámok segítségével;
- a normális eloszlástól vett távolsággal;
- további mérőszámokkal.

2.1. Kurtózis mérése a momentumokkal

A kurtózis fogalma Pearsontól származik (Pearson [1905]), aki a fogalom bevezetésével egy időben mutatót is javasolt mérésére. Ez az ún. α_4 (azaz a negyedik) mutató a második és negyedik centrális momentum segítségével jellemzi a kurtózist:

$$\alpha_4 = \frac{E\left((Y - \mu)^4\right)}{\left(E(Y - \mu)^2\right)^2} = \frac{M_4(\mu)}{M_2^2(\mu)} = \frac{M_4(\mu)}{\text{Var}^2(Y)}. \quad /1/$$

Az /1/ mutatónak az a lényege, hogy a számlálóban megjelenő negyedik centrális momentum erősen kiemeli a várható értéktől nagyban eltérő értékek szerepét, így, amennyiben viszonylag sok kis, illetve nagy érték jelenik meg az eloszlásban (széles szárny – heavy tail), a mutató értéke megnő (iránytól függetlenül). A nevezőben megjelenő variancia négyzete a különböző varianciájú eloszlások összevetését segíti, illetőleg a számláló nagyságrendjéhez igazítja a mutatót. Analitikusan belátható, hogy a központi szerepet játszó normális eloszlás esetén a mutató értéke 3, így Pearson megfogalmazásában, ha $\alpha_4 > 3$ leptokurtikus, ha $\alpha_4 = 3$ mezokurtikus, ha pedig $\alpha_4 < 3$ akkor platikurtikus eloszlásról beszélünk. (A magyar nyelvben, ennek megfelelően, a *csúcsos*, *közepes* és *lapult eloszlások* kifejezések terjedtek el, amit – mint említettük – félreérthetőségük okán ebben a tanulmányban kerülünk.)

Miután a normális eloszlás ez esetben is valamiféle etalont jelent, gyakran használják az α_4 mutató egy egyszerű transzformáltját:

$$\alpha_4^* = \alpha_4 - 3,$$

amit a többletkurtózis (excess kurtosis) mérőszámának is nevezhetünk, és értéke a normális eloszlás esetén 0, attól eltérő esetekben pedig pozitív, illetve negatív.

A Pearson-féle kurtózis mutatók nagyon elterjedtek, szinte egyeduralkodóvá váltak a leíró statisztikában: a tankönyvek szinte csak ezt említik (gyakran magát a kurtózis fogalmát is azonosítva ezekkel a mutatókkal), és minden statisztikai programcsomag részét képezik. Előnyös tulajdonságuk az, hogy elméleti és empirikus esetben egyaránt kényelmesen, jól számíthatók, kellően, de nem túlságosan érzékenyek, jól beilleszthetők a korábbi leíró mutatók (várható érték, variancia, ferdeségmutatók) rendszerébe, és mintavételi tulajdonságaik is kezelhetők. Ugyanakkor kevésbé kedvező tulajdonságokkal is rendelkeznek. Egyfelől nem mindig számíthatók, hiszen vannak fontos eloszlások, amelyeknek bizonyos momentumai nem léteznek. Emellett nincs olyan közvetlen, jól magyarázható tárgyi értelmezésük, amelyet a kvantiliseken alapuló mutatók esetén lehet látni. Végül félrevezető lehetnek azért, mert valójában nevükkel némileg ellentétben nem a csúcosságot vagy a lapultságot mérik, hanem azok valamilyen, nehezen megfogható megjelenését. (Ezzel kapcsolatban érdemes arra utalni, hogy széles (főleg alkalmazói) körben elterjedt az a nézet, miszerint a különböző paraméterű normális eloszlások csúcossága különböző. Ez persze igaz is – lásd az 1. ábrát – ám a Pearson-mutatók valamennyi normális eloszlás esetében azonosak. Ez természetesen azért van így, mert ezek a mutatók nem magát a csúcosságot, a görbe csúcspont körüli meredekségét, hanem egy ennél összetettebb, nehezebben megragadható fogalmat, a kurtózist mérik.)

Mindezek ellenére, mivel a Pearson-mutatók a kurtózis mérésének leggyakrabban használt, legelterjedtebb és megkerülhetetlen eszközei, a 3. táblázatban összefoglaljuk a legfontosabb eloszlások α_4 mutatóit.

3. táblázat

Az α_4 mutató különböző nevezetes eloszlásokon

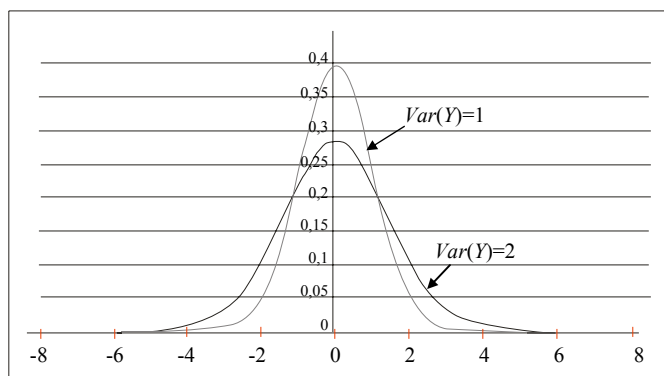
Eloszlás	α_4
Egyenletes	1,8*
Háromszög	2,4*
Normális	3*
Logisztikus	4,2*
Student- <i>t</i>	$\frac{6}{\nu - 4} + 3^{**}$
$\nu = 1$	Nem létezik
$\nu = 2$	Nem létezik
$\nu = 4$	Nem létezik
$\nu = 10$	4,0
$\nu = 100$	3,06
Lognormális	38
Exponenciális	9*
Khi-négyzet	$3 \left(1 + \frac{4}{\nu} \right)$
$\nu = 1$	15
$\nu = 2$	9
$\nu = 4$	6
$\nu = 10$	4,2
$\nu = 100$	3,12
Binomiális	$3 + \frac{1 - 6P(1 - P)}{nP(1 - P)}$
$P = 0,1$	$3 + \frac{46}{9n}$
$P = 0,5$	$3 - \frac{2}{n}$
$P = 0,8$	$3 + \frac{4}{16n}$
Poisson	$3 + \frac{1}{\sqrt{\lambda}}$

* Független a paramétereiktől.

** Ha $\nu > 4$.

Forrás: Canavos [1984], www.wikipedia.org és saját számítások.

1. ábra. Különböző varianciájú normális eloszlások sűrűségfüggvénye



Megjegyzés. A két eloszlás csúcossága láthatóan eltérő, α_4 mutatóik viszont megegyeznek.

2.2. Kvantiliseken alapuló mérőszámok

A kurtózis mérésére gyakran alkalmaznak kvantiliseken alapuló mérőszámokat is. Ezek alapötlete az, hogy minél távolabb állnak egymástól (és az eloszlás közepétől) a szélső kvantilisek (például decilisek), és mennél közelebb állnak egymáshoz (és az eloszlás közepéhez) a középső kvantilisek (például kvartilisek), annál inkább igaz az, hogy az eloszlás közepén tömörülnek az értékek, az eloszlás kicsúcsosodik, és ezzel együtt a szárnyak megnyúlnak, azaz nagy lesz a szélsőséges értékek előfordulásának valószínűsége.

A szakemberek különböző kvantilisek felhasználásával konstruáltak kurtózis mutatókat. Ezek közül itt hármat mutatunk be. Sachs [1982] K mutatóját a kvartilisek és a decilisek segítségével írta fel az alábbi formában:

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

ahol Q_1 és Q_3 az alsó és felső kvartilist, D_1 és D_9 pedig az első és utolsó decilist jelöli. A /2/ mutatót felírjuk a percentilisek (P) segítségével is:

$$K = \frac{P_{75} - P_{25}}{2(P_{90} - P_{10})}. \quad /2a/$$

A mutató konstrukciójából látható, hogy a mutató nagyobb értékei kisebb kurtózist jelölnek, hiszen, ha a számlálóban megjelenő kvantilisek távolabb állnak

egymástól, az azt jelenti, hogy kevesebb érték kap helyet az eloszlás közepén, azaz az eloszlás lapultabb, kisebb kurtózisú lesz. Másként úgy is mondhatjuk, hogy a K mutató a korábban tárgyalt α_4 -hez képest fordított jellegű mérőszám.

Némileg hasonló szellemű mutatót javasol *Moors* is (idézi *Seier* [2005]). Ő az oktiliseket használja fel mutatója képzésére:

$$M = \frac{(O_7 - O_5) + (O_3 - O_1)}{O_6 - O_2} = \frac{(P_{87,5} - P_{62,5}) + (P_{37,5} - P_{12,5})}{P_{75} - P_{25}}. \quad /3/$$

Összevetve ezt a /2/ és /2a/ mutatókkal, több különbséget, ugyanakkor több hasonlóságot is láthatunk. Valójában két lényeges különbség van közöttük: az első az, hogy az M mutató az eloszlás „szárnyközepeit” viszonyítja a középterjedelemhez, míg a K mutató az interdecilis terjedelmet. A másik lényegi különbség az, hogy az M mutató a K -hoz képest fordított irányban működik: nagyobb értékei nagyobb kurtózisra utalnak. Érdeemes megemlíteni, hogy az M mutató árnyaltabb, mivel a számlálójában az eloszlás jobb és bal oldalát külön vizsgálja, ami szimmetrikus esetben persze egyszerűsödik. Mivel az M mutató szűkebb percentilis tartományt vesz figyelembe, valamivel jobban koncentrálna az eloszlás közepére, így inkább a közép kicsúcsosodására, semmint a szárnyak tulajdonságaira érzékeny.

Az előbbieknél általánosabb kvantilis mutatót készített *Groeneveld* (idézi *Seier* [2005]), aki az eloszlásfüggvény inverzével és egy p paraméter segítségével *szimmetrikus eloszlásokra* a következő mutatót konstruálta:

$$\gamma(p) = \frac{F^{-1}(1 - (p/2)) + F^{-1}((1 + p)/2) - 2F^{-1}(0,75)}{F^{-1}(1 - (p/2)) - F^{-1}((1 + p)/2)}, \quad 0 < p < 0,5. \quad /4/$$

A /4/ mutatót $p = 0,05$ esetére konkretizálva a következőt kapjuk:

$$\gamma(0,05) = \frac{P_{97,5} + P_{52,5} - 2P_{75}}{P_{97,5} - P_{52,5}} = \frac{(P_{97,5} - P_{75}) - (P_{75} - P_{52,5})}{P_{97,5} - P_{52,5}}.$$

Ez a mutató úgy értelmezhető, hogy amennyiben az eloszlás szárnya elnyúló, a /4/ számlálójának első tagja viszonylag nagy, a második kicsi, így a hosszú szárnyat (és ezzel együtt a nagyobb csúcsosságot) a mutató nagyobb értékekkel mutatja. Ez a mutató is egyenes abban az értelemben, hogy a nagyobb kurtózist nagyobb értékekkel jellemzi, ugyanakkor p értékének változtatásával rugalmasan alakítható: kisebb p választásával inkább a eloszlásszárnyak elnyúlását, nagyobb p esetén a közép kicsúcsosodását jellemzi jobban.

Nyilvánvaló (és ezt az utolsónak tárgyalt γ mutató jól bizonyítja), hogy a kvantilisekre alapozva szinte korlátlan számú mutató képezhető. Ezek a kurtózis más és más oldalát emelik ki más és más súllyal. Közös tulajdonságuk az, hogy lényegileg minden esetben számíthatók (hiszen mind elméleti, mind empirikus eloszlások esetén csak az eloszlásfüggvényt használják fel), ugyanakkor diszkrét eloszlások esetében (nem folytonos eloszlásfüggvény esetében) igen pontatlanok lehetnek, hiszen bizonyos kvantilisok csak becslés (interpoláció) segítségével határozhatók meg. Ugyancsak kedvezőtlen tulajdonsága ezeknek a mutatóknak az, hogy mintavételi tulajdonságaik kevésbé ismertek, így becslési-hipotézisvizsgálati felhasználásuk erősen korlátozott.

A 4. táblázatban összefoglaljuk az itt bemutatott kvantiliseken alapuló mutatók számított értékeit a fontosabb eloszlásokon.

4. táblázat

*A kurtózis kvantilis mutatóinak értékei a fontosabb folytonos eloszlásokon**

Eloszlás	K	M	$\gamma(0,05)$
Egyenletes	0,312	1	0
Háromszög	0,265	1,250	0,289
Normális	0,263	1,233	0,355
Logisztikus	0,250	1,306	0,440
Student			
$\nu = 1$	0,163	1,599	0,868
$\nu = 2$	0,216	1,483	0,648
$\nu = 4$	0,242	1,401	0,504
$\nu = 10$	0,255	1,277	0,413
$\nu = 100$	0,261	1,248	0,363
Lognormális	0,185	1,791	Nem értelmezhető
Exponenciális	0,250	1,306	Nem értelmezhető
Khi-négyzet			
$\nu = 1$	0,227	1,457	Nem értelmezhető
$\nu = 2$	0,250	1,306	Nem értelmezhető
$\nu = 4$	0,258	1,262	Nem értelmezhető
$\nu = 10$	0,261	1,243	Nem értelmezhető
$\nu = 100$	0,263	1,234	Nem értelmezhető

* A számításokat Excel segítségével végeztük, az eloszlásfüggvények értékeit közelítően, mechanikus integrálással határoztuk meg.

Forrás: saját számítások.

A táblázat számait vizsgálva látható, hogy az egyes mutatók eltérő mértékkel bár, de nagyjából hasonló tendenciát mutatnak: az egyenletes eloszlás kisebb, a logisztikus, az exponenciális és főként a lognormális eloszlások nagyobb kurtózással rendelkeznek, mint a normális eloszlás. Mind a Student-féle t -, mind pedig a khinégyszet-eloszlás jól mutatja, hogy a kis szabadságfokok esetén nagy kurtózisú eloszlások, a szabadságfok növekedésével, a kurtózismutatók alapján tekintve, miként közelítenek, simulnak bele a normális határeloszlásba. Természetesen az egyes mutatók más skálákon mérik a kurtózist, ezért kérdéses lehet, hogy a számértékek nagysága vagy csak sorrendje az, ami lényeges, azaz érdemes-e magas szintű mérési skálát választani, vagy célszerűbb ordinális skálán mérni ezt a tulajdonságot. Erre a kérdésre a 3. fejezetben visszatérünk.

2.3. A szórás és az átlagos abszolút eltérés összevetésén alapuló mutatószám

Ez a viszonylag kevésbé ismert és használt mutatószám egy eloszlás szórásának és átlagos abszolút eltéréseinek összevetésére épül. Alapötlete a Geary [1936] által készített normalitási teszt volt, ahol a szerző e két mutató hányadosának tulajdonságait elemzi, és használja fel egy eloszlás normalitásának vizsgálatára és tesztelésére.

Geary erre a célra a $G = \frac{\hat{\tau}}{\hat{\sigma}}$ mutatót használja, amely normális eloszlás esetében 0,78 körül ingadozik. Meghatározva az ettől való eltérések valószínűségeit, Geary táblázatokat közölt a próba végrehajtásához.

A kurtózis méréséhez e két mutató (szórás és az átlagos abszolút eltérés) úgy használható fel, hogy a szórás jóval érzékenyebb a kiugró értékekre, így a hosszabb (szélesebb) szárnyakra is. Az ún. g -kurtózis mutatót úgy definiálták, hogy ezekből az alapokból építkezve, hasonlóan viselkedjen, mint a Pearson-féle mutató: normális eloszlás esetére eredményezzen 3 értéket és nagyobb értékei nagyobb kurtózisra utaljanak. Ennek megfelelően a g -kurtózis mutatója:

$$g = 13,29(\ln(\sigma) - \ln(\tau)), \quad /5/$$

ahol σ az eloszlás szórását, τ pedig átlagos abszolút eltérését jelöli.

Az /5/ g -kurtózis mutató előnyös tulajdonsága az, hogy minden olyan eloszlásra, melynek van szórása, könnyen számítható, viszonylag egyszerű, és közvetlenül kapcsolható az említett normalitási teszthez. Szimmetrikus és nonszimmetrikus eloszlások esetén egyaránt alkalmazható. Ugyanakkor hátrányos tulajdonsága, hogy kevésbé érzékeny, így főleg empirikus esetekben nehezen tud különbséget tenni az egyes eloszlások kurtózisa között. Az 5. táblázat a fontosabb eloszlások g -mutatóját tartalmazza.

5. táblázat

*Fontosabb eloszlások g-kurtózis mutatói**

Eloszlás	g-mutató
Egyenletes	1,91
Háromszög	2,69
Normális	3,00
Logisztikus	3,51
Student	
$\nu = 1$	Nem létezik
$\nu = 2$	Nem létezik
$\nu = 4$	4,47**
$\nu = 10$	3,42
$\nu = 100$	3,05
Lognormális	5,87
Exponenciális	4,07
Khi-négyzet	
$\nu = 1$	8,73**
$\nu = 2$	4,96**
$\nu = 4$	3,56**
$\nu = 10$	3,22
$\nu = 100$	3,03
Binomiális	
$P = 0,1$	2,99
$P = 0,5$	3,42
$P = 0,8$	2,68
Poisson	3,55

* A számításokat Excel segítségével végeztük, az eloszlásfüggvények értékeit közelítően, mechanikus integrálással határoztuk meg.

** A numerikus számítás pontatlan, ezért a mutató értékével óvatosan kell bánni.

2.4. Információelméleti mérőszámok

Az információelméleti mérőszámok közül elsősorban az entrópia az, amelyiket a leíró statisztika előszeretettel alkalmaz eloszlások egyenlőttségének jellemzésére. Ezen túlmenően a kapcsolat szorosságának jellemzésére olykor az I -divergencia mérőszámát is használják.

A kurtózis mérésére tudomásom szerint eddig nem alkalmaztak információelméleti mérőszámot. Javaslatom az, hogy kiegészítő jelleggel, *más mutatók mellett*,

használjuk az entrópia mutatóját a kurtózis jellemzésére is. Ennek indoka elsősorban a következő lehet: az entrópia közismerten olyan mutatószám, amely a rendszer (eloszlás) egyenlőtlenségét jellemzi: minél nagyobb az eloszlás egyenlősége (lapultsága) annál nagyobb az entrópia értéke. Erősen egyenlőtelen (kicsúcsosodó) eloszlások esetén az entrópia erősen lecsökken, így az entrópia a kurtózis egyfajta fordított mutatójaként is használható: minél kisebb az entrópia, annál nagyobb kurtózist jelöl.

Ennek a mutatónak az alkalmazása legalább két előnnyel járhatna: egyfelől egy olyan mutatót használnánk, amelynek tulajdonságai közismertek, más alkalmazások kapcsán gyakran amúgy is kiszámítjuk. Itt elegendő arra utalni, hogy a koncentráció elemzésénél milyen fontos szerepet tölt be az információelméleti bázison nyugvó redundancia mutató. Másfelől ez a mutató valóban csúcsosságot mér, szemben a hagyományos kurtózis mutatókkal. Amennyiben a 1. fejezetben közölt definíciókból emlékszünk azokra, amelyek a módusz körüli meredekséget emelik ki, mint a kurtózis jellemzőjét, akkor ezt a tulajdonságot leginkább az entrópia fejezi ki.

Látni kell természetesen az entrópia mutató hátrányos oldalait is. Ezek közül az első az, hogy más mutatókkal ellentétben a folytonos és a diszkrét esetek (és ennek megfelelően az elméleti és az empirikus eloszlások) nem kezelhetők teljesen analóg módon. Ha ugyanis az entrópiát diszkrét esetben értelmezzük, akkor az

$$E = -\sum p_k \ln p_k \quad /6a/$$

ismert forma jól kezelhető, hiszen a $0 \leq p_k \leq 1$ követelmény biztosítja, hogy a mutató megfelelően meghatározott intervallumban vegyen fel értékeket, így nagysága az intervallum határainak függvényében értékelhető legyen. Folytonos (elméleti) esetben az ennek megfelelő ún. *differenciális entrópia* definíciója

$$D = -\int f(y) \ln f(y) dy, \quad /6b/$$

ahol ilyen természetes és jól kezelhető határok általában nem léteznek: a D mutató értéke akár negatív is lehet. A másik, talán még ennél is komolyabb hátránya ennek a mutatónak az, hogy értéke *függ az eloszlás varianciájától*. Mivel a kurtózis eredeti definíciója kimondja az azonos varianciákat, ez kritikusnak tűnik. Az entrópia mutatója mindazonáltal nem tűnik haszontalannak. Bár nem is a szűkebb értelemben vett kurtózist, de a csúcsosságot vagy lapultságot valóban jól jellemzi. Amennyiben valódi összehasonlítást akarunk végezni ezekkel a mutatókkal, két lehetőségünk van: vagy eleve azonos varianciájú eloszlásokat választunk, és akkor az E vagy a D jól használható kurtózis mutató is lehet, vagy ha ez valami oknál fogva nem járható út, akkor az E vagy a D mutatót normáljuk a megfelelő szórással.

Mi a 6. táblázatban a D mutató számítását és értékeit mutatjuk be néhány azonos varianciájú eloszlás esetére.⁸

A táblázat alapján látható, hogy a nagy kurtózissal rendelkező eloszlások (exponenciális, lognormális) entrópiája kicsi, azaz információtartalmuk nagy-csúcsosságuk is nagy. Érdekes és továbbgondolandó ebből a szempontból az egyenletes eloszlás viszonylag nagy (a normálisnál nagyobb) csúcsossága, ami alighanem annak tudható be, hogy véges intervallumon értelmezett eloszlásról lévén szó, a szárnyai igen rövidek (végesek). Valószínűleg ez a tulajdonság tükröződik a viszonylag kis bizonytalanságot jelző alacsony entrópiában. Ez a kis példaszerű összehasonlítás már önmagában is elegendő arra, hogy elgondolkozzunk a kurtózis ellentmondásos fogalmán, és ennek következtében megpróbáljuk a jelenséget párhuzamosan minél több mutatószámmal jellemezni.

6. táblázat

Fontosabb eloszlások differenciális entrópiája

Eloszlás	D	Számértéke*
Egyenletes	$\ln(\beta - a)$	1,589
Háromszög	$1/2 + \ln\left(\frac{\beta - a}{2}\right)$	1,741
Normális	$\ln(\sigma\sqrt{2\pi e})$	1,7625
Logisztikus	$\ln(s) + 2$	1,7511
Exponenciális	$1 + \ln(\theta)$	1,3466
Lognormális	$1/2 + 0,5 \ln(2\pi\sigma^2) + \mu$	1,052

* A számítások során az eloszlások 2. táblázatban bemutatott paramétereit alkalmaztuk.

Forrás: www.wikipedia.org és saját számítások.

2.5. A normális eloszlástól vett távolság mutatója

Mint azt az előző fejezetben láttuk, a kurtózis különféle oldalait különböző mutatókkal lehet jellemezni. A kurtózis definíciójának fontos eleme a normális eloszláshoz való viszony, pontosabban az attól való eltérés (...*departure from the normal distribution*...). Ezért megkísérelhető olyan mutatószám kialakítása is, amely egysze-

⁸ A mutatók számítása esetenként meglehetősen körülményes. Ezért, bár zárt alakban kifejezhetők, itt eltekintünk néhány fontosabb eloszlás D mutatójának közlésétől. Az érdeklődőknek ezzel kapcsolatban a wikipedia.org internetes lexikon megfelelő oldalait ajánljuk tanulmányozásra.

rően az adott eloszlásnak az azonos szóródású normális eloszlástól vett távolságát méri.

A távolság több lehetséges mérőszám közül csak példaként a Pearson-féle χ^2 távolságot vettük alapul mégpedig a sűrűségfüggvényekre alkalmazva, amit a normalitásvizsgálatoknál (és függetlenségvizsgálatnál) a leíró statisztika gyakran alkalmaz:

$$d = \sum_i \frac{(f_i - f_i^*)^2}{f_i^*}, \quad /7/$$

ahol az f_i^* kitüntetett gyakoriság az alapul vett normális eloszlás esetén várt gyakoriságot, az f_i gyakoriság pedig a különböző eloszlások esetén tapasztalt megfelelő gyakoriságot jelöli.⁹ A /7/ távolságok természetesen nem tekinthetők kurtózis mérőszámoknak, már csak azért sem, mivel irányt nem mutatnak, így nem alkalmasak arra, hogy az eloszlás leptokurtikus vagy platikurtikus jellegét mutassák, de alkalmasak arra, hogy a normális eloszláshoz való hasonlatosságot jellemezzék. A 7. táblázatban és a 2. ábrán néhány gyakori, szimmetrikus és azonos varianciájú ($Var(Y) = 2$) eloszlást mutatunk be a normális eloszlástól vett távolságaikkal.

A 7. táblázat számai jól mutatják, hogy a hasonló alak önmagában még nem elegendő a kis távolságokhoz, ugyanakkor arra is rámutatnak, hogy különböző alakú sűrűségfüggvények (lásd 2. ábrát) is empirikusan igen jól közelíthetik egymást (háromszög – normális).

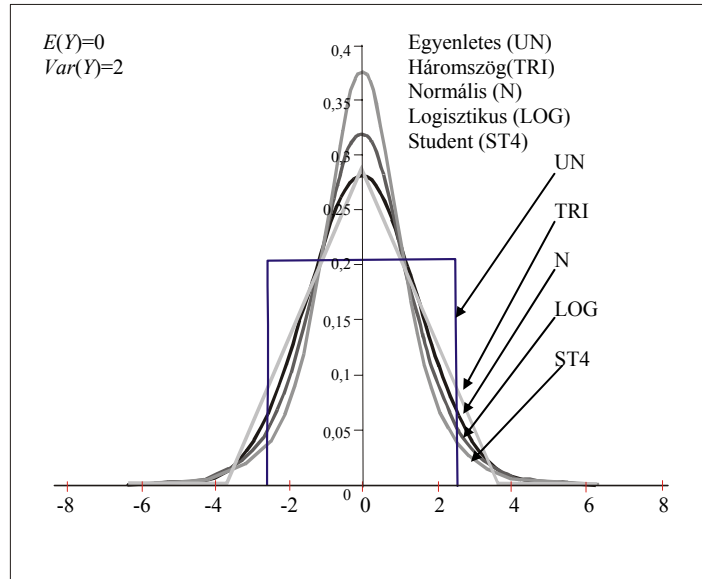
7. táblázat

<i>Szimmetrikus eloszlások távolsága a normális eloszlástól</i>	
Eloszlás	Távolság (d)
Egyenletes	3,280
Háromszög	0,333
Normális	0
Logisztikus	0,324
Student4	1,137

Forrás: saját számítások.

⁹ Itt valójában, folytonos eloszlásokról lévén szó, integrált kellene használni, de ez nagyon bonyolítaná a számításokat, ezért a diszkrét pontokon mért összehasonlítást használtuk, ami ebben az esetben elfogadható közelítést jelent.

2. ábra. Szimmetrikus, azonos varianciájú eloszlások sűrűségfüggvénye



2.6. Egyéb mérőszámok használata

A szakirodalomban a kurtózis több más szemléletű mutatószámát is megtalálhatjuk. Ezek az eloszlások különböző *transzformációi* segítségével igyekeznek kiemelni a kurtózis fontos jellemzőit. Közülük megemlítendő a Ruppert-féle hatásfüggvényre (influence function) alapozott közelítés (Ruppert [1987]), valamint a Seier és Bonett által javasolt transzformációs módszer (Seier [2005]). Ezek a módszerek, illetőleg mutatószámok a kurtózis jelenségének további árnyaltabb megközelítését teszik lehetővé, azonban eszköztáruk olyannyira eltér a leíró statisztika szokásos módszereitől, hogy itt nincs lehetőség részletezni őket. Az érdeklődők a szakirodalomból tájékozódhatnak ezek felől.

3. A különböző mérőszámok összehasonlítása

A kurtózis mutatók – mint láttuk – a jelenség más és más oldalát emelik ki, ezért eltérő képet mutatnak a különböző eloszlásokról. Ezért természetesen nem is várható

az, hogy azonos tendenciát mutassanak. Mégis érdekes – éppen a jelenség, a kurtózis jellemzésére – összehasonlítást tenni a különböző mutatók között.

Az összehasonlítást sok tényező nehezíti. Ezek között kell említeni azt, hogy nem minden mutató számítható ki minden eloszlás esetére, hogy egyes mutatók bizonyos esetekben csak nagyon pontatlanul számíthatók, valamint azt, hogy önkényes az összehasonlítás alapjául szolgáló eloszlások (és azok paramétereinek) megválasztása. Mindezen nehézségek ellenére érdekes és tanulságos az összevetés. Seier [2005] tanulmányában a szimmetrikus eloszlások esetén 18 különböző (pontosabban különbözően paraméterezett) eloszlásra és 5 mutatószámra számított a mutatók között lineáris korrelációs együtthatókat, mely összehasonlítás eredményét a 8. táblázatban idézzük fel.

8. táblázat

Kurtózis mutatók korrelációs mátrixa 18 szimmetrikus eloszlás alapján számítva

Mutató	α_4	$\gamma(0,05)$	L	SK_1	SK_2
α_4	1				
$\gamma(0,05)$	0,244	1			
L	0,389	0,917	1		
SK_1	0,438	0,879	0,992	1	
SK_2	0,515	0,797	0,966	0,985	1

Megjegyzés. Az összehasonlításban szereplő L mutató egy itt nem tárgyalt kvantilis alapú mérőszám, SK_1 és SK_2 pedig a Seier–Bonett-féle transzformációs módszerrel készült két mutató.

Forrás: Seier [2005].

A táblázat számai jól mutatják, hogy a különböző mutatók egy sor fontos eloszláson gyengén korrelálnak, leszámítva talán a két rokon mutató szoros kapcsolatát. Különösen feltűnő az α_4 gyenge korrelációja az itt vizsgált egyéb mérőszámokkal. Ez, a már említett szükségszerű különbözőség mellett felveti annak gondolatát, hogy talán nem is célszerű magas mérési skálán jellemezni az egyes eloszlások kurtózisát, hanem inkább alacsonyabb mérési skálán kell végezni az összehasonlítást, azaz csak rangsorolni az eloszlásokat a kurtózis nagysága szerint. Ezért megkíséreltük Seier számításait úgy rekonstruálni, hogy a mutatószámoknak a vizsgált eloszlásokon csak a sorrendjét vettük figyelembe, és így az egyes eloszlások között „csak” rangkorrelációt számítani. Az így kapott rangkorrelációs mátrixot a 9. táblázat mutatja.

9. táblázat

Kurtózis mutatók rangkorrelációi 18 szimmetrikus eloszlás alapján számítva

Mutató	α_4	$\gamma(0,05)$	L	SK_1	SK_2
α_4	1				
$\gamma(0,05)$	0,5666	1			
L	0,7043	0,9159	1		
SK_1	0,7420	0,8700	0,9819	1	
SK_2	0,8101	0,9572	0,9572	0,9856	1

Forrás: Seier [2005] és saját számítások.

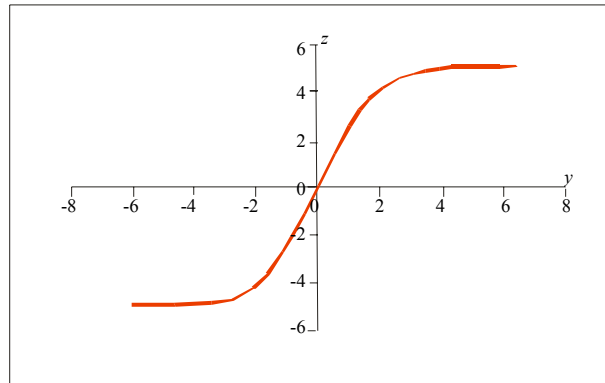
A 9. táblázat eredményeit összehasonlítva a 8. táblázatban szereplő számokkal azt tapasztaljuk, hogy az egyes vizsgált eloszlások sorrendje a különböző mutatók tükrében jóval stabilabb, megbízhatóbb, mint az eloszlásokra számított mutatók értéke. Ez, bár nem bizonyító erejű, arra utal, hogy megalapozottabb összehasonlítás tehető az egyes eloszlások között, ha a mutatókat nem próbáljuk meg magas mérési szintű skálán számítani.

Lényegileg hasonló megfontolásokból kiindulva egyes kutatók arra a véleményre jutottak, hogy nem is célszerű a hagyományos kurtózis mutatók számítása, hanem csak rangsorolni kell az eloszlásokat kurtózisaik szerint. Már az 1940-es években vizsgálták annak feltételét, hogy mikor lehet két eloszlást kurtózisuk nagysága szerint összehasonlítani, de Van Zwet volt az, aki 1964-ben módszert adott arra, hogy miként lehet eloszlásfüggvényeik alapján kurtózis szerint rendezni az egyes eloszlásokat. Van Zwet [1964] állítása az, hogy ha van két $F(y)$ -nal és $G(y)$ -nal jellemezhető szimmetrikus eloszlás, akkor annak, hogy az F -fel jelölt eloszlás kurtózisa kisebb legyen, mint a G jelű eloszlásé, szükséges és elégséges feltétele az, hogy a $z = G^{-1}(F(y))$ függvény minden $y > A$ esetén (ahol A a közös szimmetriapont) *konvex* legyen.

Erre a rangsorolásra két egyszerű példát mutatunk be. Az első példában a normális eloszlást hasonlítjuk az azonos szórású egyenletes eloszlással. Legyen tehát F a normális, G pedig az egyenletes eloszlás eloszlásfüggvénye a 2. táblázatban bemutatott paraméterezéssel. A számításokat elvégezve a $z = G^{-1}(F(y))$ függvényt a 3. ábrán láthatjuk.

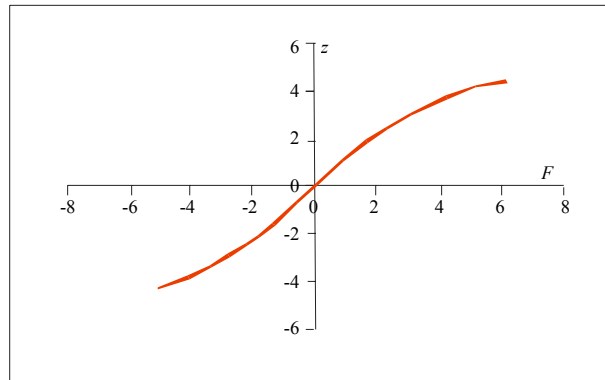
Az ábráról látható, hogy a 0 közös szimmetriaponttól jobbra a függvény konkáv, tehát az F -fel jelölt eloszlás (normális eloszlás) kurtózisa nagyobb, mint az egyenletes eloszlásé.

3. ábra. Normális és egyenletes eloszlás összehasonlítása



A másik példa a normális és a logisztikus eloszlás összevetése. Ebben az esetben (a számítások egyszerűbbé tétele érdekében) F -fel a logisztikus, G -vel a normális eloszlást jelöltük, és a kapott $z = G^{-1}(F(y))$ függvényt a 4. ábrán mutatjuk be.

4. ábra. Normális és logisztikus eloszlás összehasonlítása



Ebben az esetben a 0 ponttól jobbra a függvény jól látható módon *konkáv*, azaz az F -fel jelölt eloszlás (logisztikus eloszlás) kurtózisa nagyobb, mint a megfelelő normálisé. Ezt az eredményt erősítik meg a más mutatókkal végzett összehasonlítások is.

Az előbbi két példa alapján a vizsgált három eloszlás, kurtózisuk szerint, a következőképpen rendezhető:

$$U < N < LOG.$$

Hasonló rendezés végezhető el egy sor (de nem valamennyi) eloszlás esetére. Megjegyzendő még, hogy ez a grafikus módszer empirikus eloszlások esetére is meglehetősen kényelmesen végrehajtható.

4. Tanulságok és további kérdések

Az eddig mondottakat összefoglalva azt állíthatjuk, hogy valamelyest körbejártuk ugyan a problémát, de nem oldottuk meg: nem sikerült a mutatók között kutakodva olyat találni, amelyik minden szempontból felülmúlná, vagy pótolná a többit. Ez, a tárgy ismeretében nem is túlságosan meglepő, hiszen egy olyan fogalmat szeretnénk statisztikailag jellemezni, amelyiknek sok dimenziója van, ezért egyetlen mutatószám csak hiányosan tudja leírni. Minden itt vizsgált mutató megvilágít valamit a jelenségből (a csúcosságot, a vastag vagy hosszú szárnyakat, a görbe meredekségét, a normális eloszlástól való távolságot), de külön-külön mind csak korlátozottan. Ez, mármint, hogy egy jelenség leírására egyetlen mutató nem mindig elégséges, nem ritka a gazdasági elemzéseknél; ezért fordít egyre nagyobb figyelmet a gazdaság- és társadalomstatisztika a konzisztens indikátorrendszerek kimunkálására.

Ezért a korábban írottak fő tanulsága az, hogy alkalmazás előtt vizsgáljuk meg, valójában az előttünk álló feladat a jelenség mely oldalára utal, és annak megfelelően próbáljuk megválasztani az alkalmazandó mutatót. Ennek során természetesen figyelemmel kell lennünk arra is, hogy az egyes mutatók alkalmazási feltételei teljesüljenek. Nyilvánvaló, hogy minden kutatásnál arra törekszünk, hogy egyes jelenségeket a lehető legtömörebb formában, minél kevesebb mutatószámmal jellemezzük, de ha ez nem lehetséges vagy nem célszerű, akkor ne féljünk esetleg több mutatót használni, például a kurtózis jellemzésére. Ehhez próbál ez a tanulmány némi segítséget nyújtani.

Az elmondottakon túl még arra is felhívjuk a figyelmet, hogy a kurtózis kapcsán (is) kerüljük a téves, félreérthető fogalmazásokat! Az itteni elemzések megmutatták, hogy a nem kellően átgondolt, a korábbi eredményekből mechanikusan, néha pontatlanul átvett megfogalmazások esetenként félrevezetők lehetnek, ezért törekedjünk a jelenség minél pontosabb ismeretére és az alapján a szabatos fogalmazásokra.

Végezetül sorra kell vennünk azokat a kérdéseket, amelyek ebből, a kurtózis fogalmának és mérésének kérdéseit tárgyaló tanulmányból – részben terjedelmi okokból – kimaradtak, holott fontosságuk aligha vitatható.

Ezek közül mindenek előtt meg kell említeni azt, hogy itt csak néhány ismert és nevezetes elméleti eloszlás alapján (példáján) tárgyaltuk a kurtózis különféle kérdéseit. Kimaradt természetesen egy sor, szintén fontos ismert eloszlás esete, amelyek a

társadalmi-gazdasági elemzésekben talán kisebb szerepet játszanak. Ám mégsem ezt tekintjük fő hiánynak, hanem azt, hogy szinte egyáltalán nem vizsgáltuk az empirikus eloszlásokat. Ez annyiban érthető, hogy a vizsgált eloszlások bizonyos keretet adtak a tárgyalásnak, míg az empirikus eloszlások esetében a hasonló vizsgálatok könnyen parttalanra válhatnak. Ezekről az elmaradt vizsgálatokról itt csak annyit lehet mondani, hogy elvben ezek a mutatók empirikus eloszlásokra ugyanúgy alkalmazhatók, mint a bemutatott elméleti eloszlásokra, mégis végig kellene gondolni, hogy miként lehetne az itteni vizsgálatot kiterjeszteni. Az empirikus elemzések megtervezésekor célszerűnek látszik olyan fontos, és a gyakorlatban sokszor előforduló eloszlásokat vizsgálni, mint amilyenek a már említett jövedelemeloszlások vagy a gazdálkodó egységek méret szerinti eloszlása.

Mindenképpen jó lenne empirikus tapasztalatokat szerezni (például egy-egy jól körülhatárolt feladat elemzésén keresztül) arra vonatkozóan, hogy ezek a mutatók mennyire érzékenyek, az adathiányok, hibás adatok, a szélsőséges eloszlások hogyan befolyásolják értékeiket stb. Általában az ilyen *érzékenységvizsgálatok* nagyban segíthetik az eredmények kiterjesztését a gyakorlati problémák felé. Ehhez szorosan kapcsolódik a mutatók robusztusságának vizsgálata, ami szintén kimaradt ebből az összeállításból, holott a szakirodalomban fellelhető a robusztus kurtózis mutatók konstruálásának igénye.

Vizsgálódásaink alapvetően leíró jellegűek voltak, ezért nem foglalkoztak azzal a kérdéssel, hogy miként lehet sokasági kurtózis mutatókat mintából becsülni, illetve ezek a becslések milyen tulajdonságokkal rendelkeznek. Ez a kérdés szintén messzire vezet, ám a szakirodalom, ha nem is nagy terjedelemben, de ezzel a kérdéssel is foglalkozik. De ha már a mintából a sokaságra levonható következtetésekről beszélünk, meg kell említenünk egy másik vonatkozást, a hipotézisvizsgálatot. Az itt vizsgált mutatók ugyanis igen gyakran közvetve vagy közvetlenül kapcsolódnak ismert tesztekhez. A Pearson-féle α_4 mutató mintavételi tulajdonságainak ismeretében a normalitás vizsgálatának egyik tesztje is lehet, nem beszélve az ökonometriában népszerű Jarque–Bera-tesztről, amelyik a ferdeség és a kurtózis Pearson-féle mutatóiból épít fel próbafüggvényt a normalitás ellenőrzésére. A g -kurtózis mutatója – mint láttuk – a normalitás Geary-féle tesztjéből közvetlenül levezethető, míg a normális eloszlástól vett távolságmutató a hagyományos illesztésvizsgálatok próbafüggvénye. Ugyancsak meg kell említeni a grafikus sorbarendezési eljárást, ami jól láthatóan alapja a normalitás egyszerű P – P vagy Q – Q alapú grafikus tesztjének.

Befejezőként felmerül a kérdés, hogy miben segíthet a kurtózis fogalmának és mérésének vizsgálata az alkalmazott statisztikai kutatásokban. Két alkalmazási terület látszik fontosnak. Az egyik a normalitás/nemnormalitás kérdése, hiszen ismeretes, hogy a statisztikában a megfelelő módszerek kiválasztásakor ez döntő súllyal esik latba. A kurtózisvizsgálatok, és az ezeken alapuló tesztek ennek kapcsán rámutathat-

nak arra, hogy ebből (az egy) szempontból a normalitás tartható feltételezés-e vagy sem. A másik, a gyakorlat számára fontos kérdés talán nem is annyira a csúcosság, hanem inkább a vastag (és hosszú) szárny kérdése lehet. A vastag (hosszú) szárny ugyanis azt jelenti, hogy egy jelenség leírásakor vagy modellezésekor számolni kell azzal, hogy szélsőséges események viszonylag nagy valószínűséggel fordulnak elő, ami a leírás/modellezés eszköztárát alapvetően befolyásolja. Erre jó példát mutat *Palágyi* [2003] egy pénzügyi probléma vizsgálata kapcsán.

Irodalom

- CANAVOS, G. C. [1984]: *Applied Probability and Statistical Methods*. Little, Brown and Company. Boston–Toronto.
- GEARY, R. C. [1936]: Moments of the mean deviation to the standard deviation... *Biometrika*. 28. évf. 3–4. sz. 295–307. old.
- HUNYADI L. [2001]: *Statisztikai következtetésemélet közgazdászoknak*. Központi Statisztikai Hivatal. Budapest.
- HUNYADI L. – VITA L. [2008]: *Statisztika I-II*. AULA Kiadó. Budapest.
- KERÉKGYÁRTÓ GY. ET AL. [2008]: *Statisztikai módszerek és alkalmazásuk a gazdasági és társadalmi elemzésekben*. AULA Kiadó. Budapest.
- KOTZ, S. – JOHNSON, N. L. (szerk.) [1985]: *Encyclopedia of Statistical Sciences*. J. Wiley & Sons Inc. New York – Chicester – Brisbane – Toronto – Singapore.
- MOORS, J. J. A. [1986]: The Measuring of Kurtosis. *The American Statistician*. 40. évf. 4. sz. 283–284. old.
- RUPPERT, D. [1987]: What is Kurtosis? An influence function approach. *The American Statistician*. 41. évf. 1. sz. 1–5. old.
- PALÁGYI Z. [2003]: Pénzügyi idősorok elemzése a Lévy-hatvány GARCH-moddal. *Statisztikai Szemle*. 81. évf. 571–587. old.
- PEARSON, K. [1905]: Das Fehlergesetz und seine Verallgemeinerungen durch Fechner und Pearson. A Rejoinder, *Biometrika*. 4. évf. 1–2. sz. 169–212. old.
- PHAM, HOANG (szerk.) [2006]: *Handbook of Engineering Statistics*. Springer. New York–Heidelberg–Berlin.
- RAPPAI G. – PINTÉR J. [2007]: *Statisztika*. Pécsi Tudományegyetem Közgazdaságtudományi Kar. Pécs.
- SACHS, L. [1982]: *Applied Statistics. A Handbook of Techniques*. Springer. New York–Heidelberg–Berlin.
- SEIER, E. [2005]: *Celebrating 100 years of Kurtosis*. www.etsu.edu/math/seier/Kurto100years.doc
- VAN ZWET, W. R. [1964]: Convex transformations: A new approach to skewness and kurtosis. *Statistica Neerlandica*. 18. évf. 4. sz. 433–441. old.
- VARGHA A. [2000]: *Matematikai statisztika pszichológiai, nyelvészeti és biológiai alkalmazásokkal*. Pólya Kiadó. Budapest.
- www.wikipedia.org

Summary

The properties of an empirical distribution are measured by four types of indicators in descriptive statistics. These properties are the central tendency, variability, symmetry and departure from normal curve. The paper investigates the fourth type of indicators, the kurtosis.

It shows that the concept of kurtosis is complex and rather contradictory, so it is advisable to use more than one measure for a given distribution. Based on the moments, quantiles, absolute deviation and even on information theoretical considerations, the paper summarizes some approaches to construct measures of kurtosis. The properties and relations of these measures are demonstrated through different discrete and continuous probability distributions.