BÁNHALMI ÁRPÁD[1]

Mire jó a Yule-féle asszociációs együttható és mire nem?

George Udny Yule az 1910-es év környékén felismerte, hogy a lineáris korrelációs együttható és ennek különböző változatai alkalmatlanok bizonyos alacsony mérési szintű változók közti kapcsolatok leírására. Az 1890-es évek végén a minőségi ismérvek közti asszociációs kapcsolatok vizsgálata a logika tárgyterületéhez tartozott, Yule szándéka az volt, hogy ezeket a vizsgálatokat beemelje a statisztikai módszerek közé. A logikai alapoktól indulva konstruált – akkor még hiánypótló jelleggel – 2×2-es táblák esetén használható, az asszociációs kapcsolatok mértékét kifejező mutatókat: az asszociációs és a kolligációs együtthatókat. A magyar szakirodalomban – különösen a statisztikatankönyvekben – figyelmen kívül hagyják a Yule által bevezetett mutatók konstruálásánál alkalmazott alapvetéseket, így a mutatók felületes és félrevezető értelmezését adják.

Azt gondolhatnánk, hogy a 2×2-es kontingenciatáblák részletes vizsgálata napjainkra rég lezárult, de ez koránt sincs így. Az elmúlt száz évben számtalan mutatót konstruáltak alternatív ismérvek asszociációs kapcsolatának vizsgálatára, a nemzetközi szakirodalomban – kellően megalapozott és elfogadott módon – húsznál is több mutatót használnak. A nominális és ordinális szinten mérhető alternatív ismérvek kapcsolatát, az egyes mutatók értelmezési módjait és a mutatók egymással való kapcsolatát jelenleg is aktívan vizsgálják.

A gazdasági felsőoktatásban a 2×2-es táblák vizsgálatára alkalmazott mutatókat összehasonlítva azt tapasztaljuk, hogy bizonyos megoszlások esetén feltűnően különböző értéket adnak, ezért a részletes összehasonlításukkal választ keresünk az eltérés okára. A Yule-féle mutatón kívül a tananyagban és a számonkérésben a Csuprov- és a Cramer-mutató szerepel, amelyek értékei 2×2-es táblák esetén megegyeznek. Az összehasonlítást mégis az – előjellel ellátott – φ és a Yule-féle együttható esetén

1 Budapesti Gazdasági Főiskola Külkereskedelmi Kar Módszertani Intézeti Tanszéki Osztály; e-mail cím: banhalmi.arpad@kkk.bgf.hu.

végezzük el, mivel a φ abszolút értéke megegyezik a Csuprov- és Cramer-mutatók értékével, az előjele pedig ugyanaz, mint a Yule-mutatónak.

Rögzített peremgyakoriságok esetén már Yule is vizsgálta az általa konstruált asszociációs együttható és más asszociációs mérőszámok eltérését és kapcsolatát. A peremgyakoriságokra tett feltételt feloldva is lehet egzakt következtetéseket levonni a φ és a Yule-mutató eltérésének alsó határáról: a Yule-mutató abszolút értéke legalább annyi, mint a $2\varphi/(1 + \varphi^2)$ abszolút értéke, az egyenlőség csak szimmetrikus 2×2-es táblák esetén teljesül. Tehát a Yule-mutató minden esetben erősebb asszociációs kapcsolatot jelez, mint a φ – ezáltal a gazdasági felsőoktatásban tanított Csuprov- és Cramer-mutató is.

A φ és a Yule-mutató szükségképpen fennálló eltérésére a magyarázatot az adja, hogy míg a vizsgálatba vont változók függetlenségét a φ és a Yule-mutató azonos módon veszi figyelembe, a teljes asszociációt másként. Mindkét mutató konstrukciójának alapgondolata, hogy az asszociációs-disszociációs kapcsolat mértékét – a pontosan definiált negatív és pozitív asszociáció figyelembevételével – a teljes asszociáció vagy disszociáció és a függetlenség között mérje. Teljes disszociáció esetén mindkét mutató értéke –1, függetlenség esetén 0, és ha teljes asszociáció tapasztalható, 1. Részletes elemzéssel és a Yule-féle konstrukciós alapfeltevések figyelembevételével nyilvánvalóvá válik, hogy a teljes asszociáció (disszociáció) definíciója eltér a két mutató esetén. A Yule-féle meghatározás szerint teljes asszociációról vagy disszociációról akkor beszélünk, ha *legalább* egy ismérvváltozat szükségképpen együtt jár a másik ismérv valamelyik változatával, a teljes asszociáció ezen meghatározásának tesz eleget a Yule-mutató. A φ meghatározásánál a teljes asszociáció egy másik meghatározása játszik szerepet: ebben az esetben teljes asszociációról vagy disszociációról akkor beszélünk, ha *mindegyik* ismérvváltozat szükségképpen együtt jár a másik ismérv valamelyik változatával.

A címben feltett kérdésre a válasz: a Yule-mutató és a φ a teljes asszociáció mértékét más-más értelemben méri, amit e mutatók értelmezésénél és az értékeik összevetésénél figyelembe kell venni.
Kulcsszavak: asszociáció, asszociációs mérték, Yule, kontingencia, kolligáció

Bevezető

Ebben a tanulmányban a *Y*ule-féle asszociációs együttható és a φ értékeinek eltérésére keresünk magyarázatot. Nagyszámú véletlenszerűen generált 2×2-es tábla segítségével szimuláljuk az együttes eloszlásukat, majd a szimuláció alapján levont következtetéseket matematikailag igazoljuk. A két mutató eltérésére kapott alsó becslésre a mutatók eltérő definíciójában keresünk magyarázatot, és azt találjuk, hogy a két mutató esetén a teljes asszociáció fogalma eltér. Ezek alapján javaslatot teszünk a mutatók helyes használatára és értelmezésére: a Yule-mutató egyhez közeli abszolút értéke azt fejezi ki, hogy legalább, a φ pedig azt, hogy pontosan két ismérvváltozat esetén mutatható ki erős asszociáció. Bár a két mutatóból eltérő következtetést lehet levonni az együttes asszociációról, az értékeikből nyerhető információ kiegészíti egymást.

Vizsgálatunk célja, hogy tisztázzuk a két, a 2×2-es táblákra alkalmazott asszociációs mérték használatában és értelmezésében tapasztalható különbséget. A témaválasztást az indokolja, hogy a magyar szakirodalomban a *Y*ule-féle asszociációs együttható és a φ értelmezési és alkalmazási köre nem különül el, ezzel látszólagos ellentmondásba keveredhetünk, ha alternatív változatokkal rendelkező ismérvek kapcsolatát értékeljük. E bizonytalanság megszüntetése érdekében két módszerrel hasonlítjuk össze a mutatókat: egyrészt értékeik – szimuláción és matematikai becsléseken alapuló – összevetésével tisztázzuk, hogy a különböző eloszlásoknál milyen *mértékű* eltérésre számíthatunk, másrészt definíciójuk összehasonlításával a különböző eltérések *értelmezését* dolgozzuk ki. Ebben a tanulmányban a Yule-féle asszociációs együttható és a φ megfelelő értelmezésének kidolgozására vállalkozunk.

Röviden áttekintjük a két mutató viszonyáról eddig ismert összefüggéseket. George Udny Yule az 1900-as év környékén felismerte, hogy a lineáris korrelációs együttható, és ennek különböző változatai alkalmatlanok bizonyos alacsony mérési szintű változók közti kapcsolatok leírására (Yule 1900: 257). Az 1890-es évek végén a minőségi ismérvek közti asszociációs kapcsolatok vizsgálata a logika tárgyterületéhez tartozott, Yule szándéka az volt, hogy ezeket a vizsgálatokat beemelje a statisztikai módszerek közé (Yule 1900: 258). *Logikai alapoktól* indulva konstruált – akkor még hiánypótló jelleggel – 2×2-es táblák esetén használható, az asszociációs kapcsolatok mértékét kifejező mutatókat: az asszociációs és a kolligációs (Yule 1912) együtthatókat. Azt gondolhatnánk, hogy a 2×2-es kontingenciatáblák részletes vizsgálata napjainkra rég lezárult, de ez koránt sincs így. Az elmúlt száz évben számtalan mutatót konstruáltak alternatív ismérvek asszociációs kapcsolatának vizsgálatára; a nemzetközi szakirodalomban – kellően megalapozott és elfogadott módon – húsznál is több mutatót használnak (Tan et al. 2004), köztük a Q-t és a φ-t. A nominális és ordinális szinten mérhető alternatív ismérvek kapcsolatát, az egyes mutatók értelmezési módjait és a mutatók egymással való kapcsolatát jelenleg is aktívan vizsgálják (Ekström 2008), ehhez a kutatási irányhoz kapcsolódik a Q és a φ *minimális eltérésének* újszerű vizsgálata. Rögzített peremgyakoriságok esetén már Yule is vizsgálta az általa konstruált asszociációs együttható és más asszociációs mérőszámok eltérését és kapcsolatát (Yule 1912), ezt a módszert meghaladva a peremgyakoriságokra tett feltétel feloldásával egzakt eredményeket adunk a φ és a Yule-mutató eltérésének alsó határára.

A tanulmány első részében egy valós adatokon alapuló – és a pedagógiai kutatásban tipikusnak mondható – példán illusztráljuk a két mutató eltérését, ezzel hangsúlyozva, hogy a *statisztikai gyakorlat* számára vonunk le releváns következtetéseket. A példa adatai alapján kiemeljük azt a *tapasztalatot,* hogy a két mutató *látszólag* eltérő erősségű asszociációs kapcsolatot jelez. A következő részben azt vizsgáljuk, hogy e jelentős eltérés a két mutató értékében csak ritkán, elszigetelten fordul-e elő, vagy általánosnak tekinthető. Ennek érdekében végzünk egy 2000 darab véletlenszerűen generált 2×2-es táblát tartalmazó szimulációt, ami a két mutató együttes eloszlásáról érdekes információkat szolgáltat: ha a φ közepes erősségű kapcsolatot jelez, a Yule-féle asszociációs együttható jóval erősebb kapcsolatra utal, mint a φ. A szimuláció során feltárt minimális eltérés összefüggéseit matematikai módszerekkel pontosan megadjuk. A matematikai becslés alapgondolata, hogy belátjuk, a kolligációs együttható a φ és a Yule-mutató közé esik, illetve a Yule-mutató kifejezhető a kolligációs együttható függvényeként. Az esetenként szembeötlő eltérés a φ és a Yule-mutató értékei között tehát nem esetleges, az okát a mutatók definíciója segítségével kívánjuk felfedni. A tanulmány következő részében sorra vesszük a mutatók néhány lehetséges definícióját: koordinációs viszonyszámokon, függetlenség esetén fennálló gyakoriságokhoz való viszonyításon, ellentétes irányú asszociáció fogalmán és a teljes asszociáció meghatározásán alapuló bevezetését. A következő részben összegezzük a leírtakat, és megállapítjuk, hogy a megalapozási módok közül a teljes asszociáció fogalmából kiinduló definíció nyújt kellő magyarázatot a mutatók eltérésére, egyúttal világossá teszi, hogy a két mutató a vizsgált ismérvek kapcsolatáról mást-mást mond. Ezután kifejtjük, hogy a két mutató által szolgáltatott információ kiegészíti egymást, amit a használatuknál és az értelmezésüknél javasolt figyelembe venni. Végül az aktuális módszertani problémákhoz kapcsoljuk az asszociációs mértékek vizsgálatát.

A problémát illusztráló gyakorlati példa

Az asszociációs kapcsolat erősségének mérésénél felmerülő problémákat először egy konkrét – valós mérésen alapuló – példán szemléltetjük. A 2013/2014-es tanév őszi félévében a BGF KKK-n két szemináriumi csoport – a 7-es és a 11-es csoport – hallgatói körében vizsgáljuk két elemi matematikafeladat helyes megoldása közti asszociációs kapcsolatot. A két csoportba összesen 63 hallgató járt, közülük 52-en írták meg a félév eleji szintfelmérő dolgozatot. A továbbiakban levont következtetések csak a szintfelmérőt megíró 52 hallgatóra vonatkoznak. A szintfelmérő feladatsor 90 elemi matematikafeladatot tartalmazott, amelyek közül kettő:

A feladatsor, 4. feladat: „Fejezze ki 2 hatványaként a következő kifejezést! $\sqrt[3]{2^4} = \ldots$"

A feladatsor, 27. feladat: „Gyöktelenítse az adott tört nevezőjét! $\frac{4}{\sqrt{7}-\sqrt{3}} = \ldots$"

A feladatsoroknak három – *A*, *B* és *C* – változata volt, a 4. és 27. feladatok mindhárom feladatsoron a fenti típusoknak feleltek meg. A szintfelmérő során azt értékeltük, hogy az egyes feladatokat hibátlanul *teljesítette*-e a hallgató. A 4. és 27. feladat esetén a következő megoszlást kaptuk:

1. táblázat: *A hallgatók számának* (fő) *megoszlása a 4. és 27. feladat hibátlan teljesítése szerint*

		27. feladat		
		Nem teljesítette	Teljesítette	**Összesen**
4. feladat	Nem teljesítette	33	1	**34**
	Teljesítette	13	5	**18**
	Összesen	**46**	**6**	**52**

Forrás: saját szerkesztés (2014)

A két feladat teljesítése közti kapcsolat erősségére vagyunk kíváncsiak, ezért kiszámítottuk a Yule-mutatót (értéke **0,8539**) és a Csuprov-mutatót (értéke **0,3698**). A gazdasági felsőoktatásban széles körben alkalmazott *tankönyvek szerint* mindkét mutató abszolút értékének 0 és 1 közé kell esni: a 0-hoz közeli értékek gyenge, az 1-hez közeli értékek erős asszociációs kapcsolatra utalnak (Korpás 1996: 133). Esetünkben – a tankönyvi értelmezés szerint – a feladatok teljesítése között „a Yule-mutató erős, a Csuprov-mutató közepesnél gyengébb kapcsolatot jelez", ami *ellentmondónak tűnik*. Példánk szempontjából, de általában gyakorlati szempontból is fontos lenne tisztázni a következő kérdéseket: A két mutató közti eltérés milyen mértékű, mekkora eltérés adódik a különböző megoszlásokra, milyen esetekben adnak közel azonos értékeket? Ha a két mutató jelentősen eltér egymástól, hogyan értelmezhetők az eredményeik?

A kérdések megválaszolásához a két mutató nagyságrendjét kell összehasonlítani és tisztázni kell a konstrukciójuknál alkalmazott alapvetéseket, hogy azok milyen következménnyel járnak az értelmezésükre.

A φ és a Yule-mutató értékeinek összehasonlítása

Az A és B alternatív – rendre A_1, A_2 és B_1, B_2 ismérvváltozatokkal rendelkező – ismérvek esetén az együttes és peremgyakoriságokat az alábbi kombinációs tábla szemlélteti:

1. ábra: *Sematikus 2×2-es kombinációs tábla*

		B		
		B_1	B_2	**Összesen**
A	A_1	f_{11}	f_{12}	$\mathbf{f_{1+}}$
	A_2	f_{21}	f_{22}	$\mathbf{f_{2+}}$
	Összesen	$\mathbf{f_{+1}}$	$\mathbf{f_{+2}}$	**n**

Ezek alapján a Yule-mutató[2] számítási módja:

$$Q = \frac{f_{11} \cdot f_{22} - f_{12} \cdot f_{21}}{f_{11} \cdot f_{22} + f_{12} \cdot f_{21}}$$

2 A Yule-mutatót a magyarországi tankönyvekben *Y*-nal (Korpás 1996; Sándorné et al. 2013) vagy *a*-val (Köves – Párniczky 1973; Kerékgyártó et al. 2001) jelölik. Yule eredetileg *Q*-val jelölte (Yule 1900), és a külföldi szakirodalomban a mai napi Yule-féle Q együtthatóként hivatkoznak rá (Kotz 2006). Jelen tanulmányban azért is indokolt ez a jelölés, mert *Y*-nal a Yule-féle kolligációs együtthatót fogjuk jelölni (Yule – Kendall 1964).

A Csuprov- és Cramer-mutató meghatározási módja:

$$T = C = \frac{|f_{11} \cdot f_{22} - f_{12} \cdot f_{21}|}{\sqrt{f_{+1} \cdot f_{+2} \cdot f_{1+} \cdot f_{2+}}}$$

A φ mutató a T és C „előjeles változata”:

$$\varphi = \frac{f_{11} \cdot f_{22} - f_{12} \cdot f_{21}}{\sqrt{f_{+1} \cdot f_{+2} \cdot f_{1+} \cdot f_{2+}}}$$

A φ mutató abszolút értéke megegyezik a Csuprov- és Cramer-mutatók értékével, az előjele pedig azonos a Yule-mutató előjelével.

A Q értékét maga Yule is összehasonlította más asszociációs mérőszámokkal (Yule 1912), de ezt rögzített peremgyakoriságok mellett tette. Most arra törekszünk, hogy feloldjuk ezt a feltételt, a Q és a φ értékét úgy hasonlítsuk össze, hogy ne kelljen rögzített peremgyakoriságokra hivatkozni. Ennek egyik lehetséges módja, hogy nagyszámú, véletlenszerűen kitöltött 2×2-es tábla adataiból meghatározzuk a Q és a φ mutatókat, és az összetartozó értékeket összehasonlítjuk. Ezután a szimulált eloszlások alapján levont következtetéseket szigorú matematikai eszközökkel is belátjuk.

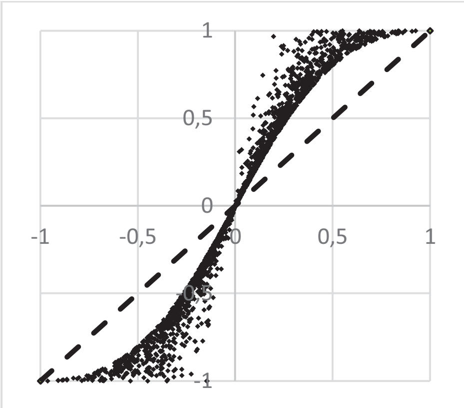
Szimuláción alapuló összehasonlítás

A szimuláción alapuló összehasonlítás az asszociációs mérőszámok elemzésének elfogadott módszere (Tan et al. 2004; Ekström 2009). Nagy tömegben, véletlenszerűen kitöltött 2×2-es táblákból állítjuk elő és hasonlítjuk össze Q és φ értékeit, majd a legegyszerűbb módszerrel, grafikus elemzéssel – aminek során a φ együttható függvényében ábrázoljuk Q értékeit – teszünk megállapításokat. Egy ilyen szimuláció végeredményét mutatja a *2. ábra,* ahol az összesen 2000 véletlenszerűen[3] generált 2×2-es tábla alapján számított (φ; Q) pontokat ábrázoltuk. A grafikonon láthatunk egy szaggatott vonalat, ami a (φ; φ) pontok helyét jelöli. A φ és Q értékei közt azokban az esetekben találunk egyezőséget, amikor valamely

3 Az egyes táblákban, a négy cellába egymástól függetlenül, az együttes gyakoriságok helyére kerültek 0-tól 1000-ig véletlensze rűen nemnegatív egész értékek.

(φ; Q) pont a szaggatott vonalra esik. Ilyen lehet a vonal két végpontja és az origó: a teljes függetlenség és a függvényszerű kapcsolat esete. A szaggatott vonaltól távolabbi pontok a két mutató nagyobb, a közelebbi pontok kisebb eltérésére utalnak. A pontfelhőt szemlélve azt a sejtést fogalmazhatjuk meg, hogy a φ és a Q eltérése nem teljesen véletlenszerű, a φ abszolút értékénél a Yule-féle Q mutató abszolút értéke – a függetlenségtől és a függvényszerű kapcsolattól eltekintve – határozottan nagyobb. A szaggatott vonalhoz legközelebbi pontok – a negatív tartományon a pontfelhő „teteje", a pozitív tartományon az „alja" – olyan rajzolatot mutatnak, amit érdemes valamilyen függvénnyel leírni. Az is megfigyelhető, hogy ez a mutatók minimális eltérését adó (függőleges) távolság a φ mutató ±0,5 körüli értékeinél a legnagyobb, körülbelül 0,3.

2. ábra: *A φ függvényében* (vízszintes tengely) *a Q mutató* (függőleges tengely)



Matematikai becsléseken és differenciálszámításon alapuló összehasonlítás

A grafikus elemzés alapján megfogalmazott sejtés egzakt módon igazolható. Yule az asszociációs együttható mellett (Q) konstruált egy hasonló tulajdonságokkal rendelkező másik mutatót, amit kolligációs együtthatónak nevezett (Y) (Yule 1912). A két mutató között a következő kapcsolat áll fenn: $Q = \frac{2Y}{1+Y^2}$ (Yule – Kendall 1964). Először belátjuk, hogy a $|\varphi| \leq |Y|$ reláció minden esetben teljesül (pontosan tisztázzuk az egyenlőség feltételét is), majd kihasználjuk azt a tényt, hogy a [–1; 1] intervallumon a $Q = \frac{2Y}{1+Y^2}$ szigorúan monoton nő Y függvényében, ahonnan a konvexitás figyelembevételével $|\varphi| \leq \frac{2|\varphi|}{1+\varphi^2} \leq \frac{2|Y|}{1+Y^2} = |Q|$ adódik. Ezzel meghatározható tetszőleges φ esetén a két mutató legkisebb eltérése, és azt is megadjuk, hogy ez a minimális eltérés milyen 2×2-es táblák esetén valósul meg.

A φ és a kolligációs együttható abszolút értékének összehasonlítása

A kolligációs együttható Yule meghatározása szerint:

$$Y = \frac{\sqrt{f_{11} \cdot f_{22}} - \sqrt{f_{12} \cdot f_{21}}}{\sqrt{f_{11} \cdot f_{22}} + \sqrt{f_{12} \cdot f_{21}}}$$

Ebből a számláló konjugáltjával való bővítés után kapjuk:

$$Y = \frac{f_{11} \cdot f_{22} - f_{12} \cdot f_{21}}{\left(\sqrt{f_{11} \cdot f_{22}} + \sqrt{f_{12} \cdot f_{21}}\right)^2}$$

Állításunk szerint $|\varphi| \leq |Y|$, tehát $\frac{|f_{11} \cdot f_{22} - f_{12} \cdot f_{21}|}{\sqrt{f_{+1} \cdot f_{+2} \cdot f_{1+} \cdot f_{2+}}} \leq \frac{|f_{11} \cdot f_{22} - f_{12} \cdot f_{21}|}{\left(\sqrt{f_{11} \cdot f_{22}} + \sqrt{f_{12} \cdot f_{21}}\right)^2}$, ami akkor és csak akkor teljesül, ha

$$\sqrt{f_{+1} \cdot f_{+2} \cdot f_{1+} \cdot f_{2+}} \geq \left(\sqrt{f_{11} \cdot f_{22}} + \sqrt{f_{12} \cdot f_{21}}\right)^2 .$$

Mivel az egyenlőtlenség mindkét oldala nemnegatív, négyzetre emeléssel ekvivalens egyenlőtlenséghez jutunk:

$$f_{+1} \cdot f_{+2} \cdot f_{1+} \cdot f_{2+} \geq \left(\sqrt{f_{11} \cdot f_{22}} + \sqrt{f_{12} \cdot f_{21}}\right)^4 .$$

A peremgyakoriságokat felírjuk a megfelelő együttes gyakoriságok összegeként:

$$(f_{11} + f_{12}) \cdot (f_{21} + f_{22}) \cdot (f_{11} + f_{21}) \cdot (f_{12} + f_{22}) \geq \left(\sqrt{f_{11} \cdot f_{22}} + \sqrt{f_{12} \cdot f_{21}}\right)^4 .$$

Az egyszerűbb áttekinthetőség érdekében bevezetjük a következő jelöléseket:

$$a = f_{11}; b = f_{12}; c = f_{21}; d = f_{22} .$$

A számtani-mértani közepek közti nevezetes összefüggés alapján tetszőleges nemnegatív (A, B) valós számpár esetén:

$$\sqrt{AB} \leq \frac{A+B}{2}$$

Mindkét oldalt szorozva 2-vel:

$$2\sqrt{AB} \leq A + B$$

Ez utóbbi egyenlőtlenséget alkalmazzuk a következő párokra:

(a^2bc, bcd^2);	(a^2cd, abd^2);	(b^2cd, abc^2)
(ab^2d, ac^2d);	(a^2bd, acd^2);	(bc^2d, ab^2c)

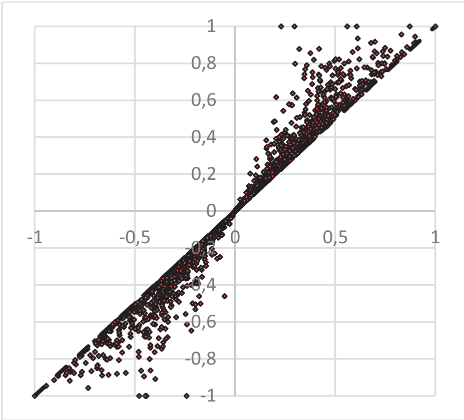
$2abcd \leq a^2bc + bcd^2$;	$2(\sqrt{ad})^3\sqrt{bc} \leq a^2cd + abd^2$;	$2\sqrt{ad}(\sqrt{bc})^3 \leq b^2cd + abc^2$
$2abcd \leq ab^2d + ac^2d$;	$2(\sqrt{ad})^3\sqrt{bc} \leq a^2bd + acd^2$;	$2\sqrt{ad}(\sqrt{bc})^3 \leq bc^2d + ab^2c$.

A hat egyenlőtlenséget összeadva, és mindkét oldalt megnövelve $(a^2d^2 + b^2c^2 + 2abcd)$-vel, a bizonyítandó állítást kapjuk: $(\sqrt{ad} + \sqrt{bc})^4 \leq (a + b)(a + c)(b + d)(c + d)$.

Figyelembe véve, hogy a peremgyakoriságok pozitívak, az egyenlőség csak szimmetrikus[4] tábla ($a = d$ és $b = c$) esetén teljesül. Ez a φ és a kolligációs együtthatóra (Y) vonatkozóan azt jelenti, hogy függetlenség, függvényszerű kapcsolat és szimmetrikus (az átlókban megegyező értékeket tartalmazó) táblák esetén egyezik meg a két mutató értéke. Azt is megállapíthatjuk, hogy ezekhez az esetekhez közeli 2×2-es táblák esetén a φ és Y értéke egymáshoz közeli, úgyhogy az Y abszolút értéke minden esetben legalább akkora, mint a φ abszolút értéke. A φ és Y nagyságrendi viszonyait szemlélteti a *3. ábra,* ami – az előbbiek során is használt – véletlenszerűen generált 2×2-es táblák adatai alapján mutatja az Y értékeit (függőleges tengely) a φ értékeinek függvényében (vízszintes tengely).

4 Itt más értelemben használjuk a „szimmetrikus” kifejezést, mint a mátrixelméletben. Szimmetrikus tábla alatt itt „középpontosan szimmetrikust” értünk, a 2×2-es tábla főátlójában és mellékátlójában is egyforma gyakoriságok szerepelnek.

3. ábra: *A φ függvényében* (vízszintes tengely) *az Y mutató* (függőleges tengely)



A φ és a Yule-féle asszociációs együttható minimális eltérése

Ha a $Q = \frac{2Y}{1+Y^2}$ összefüggést tekintjük ($-1 \leq Y \leq 1$), megállapíthatjuk, hogy a Q asszociációs együttható Y szerinti elsőrendű deriváltja $-1 < Y < 1$ esetén pozitív:

$$\frac{dQ}{dY} = \frac{2(1 - Y^2)}{(1 + Y^2)^2} > 0$$

Tehát $\frac{2Y}{1+Y^2}$ az Y függvényében szigorúan monoton nő a $]-1; 1[$ intervallumon. Kihasználva, hogy $\frac{2Y}{1+Y^2}$ az Y függvényében páratlan, az imént megállapított $|\varphi| \leq |Y|$ relációból következik:

$$\frac{2|\varphi|}{1 + \varphi^2} \leq \frac{2|Y|}{1 + Y^2}$$

Mivel $|\varphi| \leq 1$, ezért $1 + \varphi^2 \leq 2$, így $1 \leq \frac{2}{1+\varphi^2}$, ahonnan $|\varphi| \leq \frac{2|\varphi|}{1+\varphi^2}$, tehát

$$|\varphi| \leq \frac{2|\varphi|}{1 + \varphi^2} \leq \frac{2|Y|}{1 + Y^2} = |Q|$$

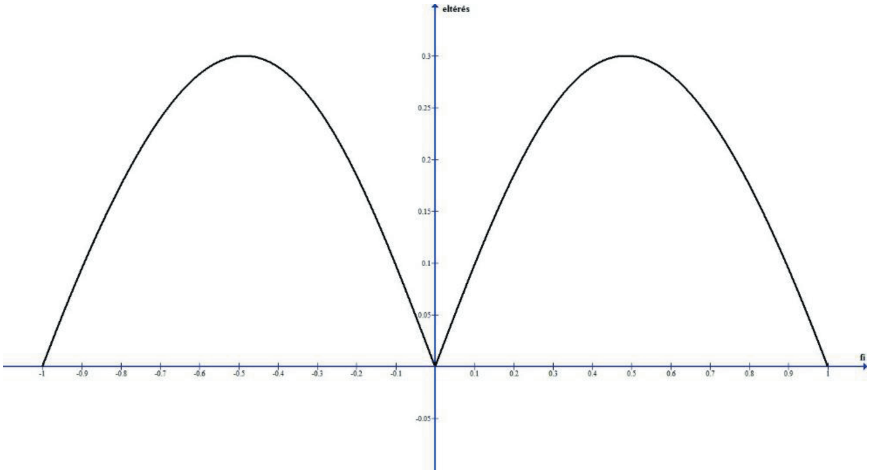
Ezzel beláttuk, hogy a φ abszolút értékénél – így a Csuprov- és Cramer-mutatónál – a Yule-féle asszociációs együttható (Q) abszolút értéke, a függetlenség és a függvényszerű kapcsolat esetét leszámítva, minden esetben nagyobb. Sőt azt is meg tudjuk mondani, hogy ez az eltérés legalább mennyivel nagyobb.

A φ és Q azonos előjelűek, ezért a minimális eltérésüket a következő reláció adja meg:

$$|\varphi - Q| \geq \left|\varphi - \frac{2\varphi}{1+\varphi^2}\right|$$

A *4. ábra* a φ és a Q minimális eltérését szemlélteti.

4. ábra: *A φ függvényében* (vízszintes tengely) *a φ és a Q mutató minimális eltérése* (függőleges tengely)



Természetesen a $\left|\varphi - \frac{2\varphi}{1+\varphi^2}\right|$ maximumhelyei és maximumértékei függvényelemzéssel pontosan meghatározhatók, a minimális eltérés a legnagyobb, ha $\varphi = \pm\sqrt{\sqrt{5}-2} = \pm\, 0{,}4859$, és a két mutató minimális eltérése ebben az esetben:

$$|\varphi - Q| \geq \frac{(3-\sqrt{5})\sqrt{\sqrt{5}-2}}{\sqrt{5}-1} = 0{,}3003$$

Az is belátható, ha φ = 0,4859, akkor a Q minden 0,4859 + 0,3003 = 0,7862 és 1 közötti értéket – alkalmas 2×2-es tábla esetén – felvehet. Ez általánosabban is igaz, tetszőleges $\varphi \neq 0, \pm 1$ esetén a Q

(1) a φ pozitív értékeire a $\left[\frac{2\varphi}{1+\varphi^2}; 1\right]$ és

(2) a φ negatív értékeire a $\left[-1; \frac{2\varphi}{1+\varphi^2}\right]$ intervallum minden elemét felveheti valamely alkalmas eloszlásra.

Ehhez fel kell használni Q azon tulajdonságát, hogy ha a 2×2-es tábla valamely sorát vagy oszlopát megszorozzuk egy pozitív számmal, a Q értéke változatlan marad (Yule 1912). Legyen adott egy pozitív φ, be fogjuk látni, hogy Q minden lehetséges értéket felvehet $\frac{2\varphi}{1+\varphi^2}$ és 1 között (negatív φ értékekre hasonló bizonyítás adható, ezért csak a pozitív esetet részletezzük). A bizonyítást két lépésben végezzük el.

I. Először belátjuk, hogy van olyan együttes eloszlás bármely $0 < \varphi < 1$ esetén, amelyre minden

$$\frac{2\varphi}{1+\varphi^2} \leq Q < 1$$

érték megvalósul. Az *5. ábrán* látható kombinációs tábla adja meg azokat az eloszlásokat, amelyekre a Q értéke legközelebb van a φ értékéhez adott φ esetén, ahol a egy pozitív arányossági tényező. Egyszerű számolással adódik, hogy ebben az esetben valóban $Q = \frac{2\varphi}{1+\varphi^2}$.

5. ábra: *Szimmetrikus 2×2-es kombinációs tábla*

		B		
		B_1	B_2	**Összesen**
A	A_1	$a(1+\varphi)$	$a(1-\varphi)$	$\mathbf{2a}$
	A_2	$a(1-\varphi)$	$a(1+\varphi)$	$\mathbf{2a}$
	Összesen	$\mathbf{2a}$	$\mathbf{2a}$	$\mathbf{4a}$

Most tekintsünk egy φ-nél nagyobb φ_0 értéket, ami az *5. ábrán* látható eloszlások esetén adódik, ha φ helyébe φ_0-t írunk, ugyanezen eloszlásokra $Q = \frac{2\varphi_0}{1+\varphi_0^2}$. Amennyiben ennek a táblának az első sorát megszorozzuk egy pozitív x együtthatóval, a Q értéke változatlan marad, a φ értéke így módosul:

$$\varphi(x) = \frac{2x\varphi_0}{\sqrt{x[(1+\varphi_0)x+1-\varphi_0][(1-\varphi_0)x+1+\varphi_0]}}$$

A függvény deriváltja:

$$\varphi'(x) = \frac{x(1-x^2)\varphi_0(1-\varphi_0{}^2)}{\left(\sqrt{x[(1+\varphi_0)x+1-\varphi_0][(1-\varphi_0)x+1+\varphi_0]}\right)^3}$$

Mivel a $\varphi(x)$ deriváltja $\varphi_0 \neq 1$ esetén a $]0;\ 1[$ intervallumon pozitív, illetve $\lim_{x\to 0+}\varphi(x) = 0$ és $\varphi(1) = \varphi_0$, a Bolzano-tétel értelmében a $\varphi(x)$ minden értéket pontosan egyszer vesz fel a $]0;\ \varphi_0]$ intervallumon ($0 < x < 1$), így valamely x_0-ra $\varphi < \varphi_0$ miatt a φ értékét is. Tehát erre az x_0 szorzóra teljesül, hogy pont az adott φ és Q érték adódik asszociációs mérőszámként az *6. ábrán* megjelenített eloszlás mellett.

6. ábra: *Olyan eloszlások, amelyekre a φ és a Q az előre adott értékeket veszik fel* (a > 0)

		B		**Összesen**
		B_1	B_2	
A	A_1	$ax_0(1+\varphi_0)$	$ax_0(1-\varphi_0)$	$\boldsymbol{2ax_0}$
	A_2	$a(1-\varphi_0)$	$a(1+\varphi_0)$	$\boldsymbol{2a}$
	Összesen	$\boldsymbol{a}[x_0(1+\varphi_0)+(1-\varphi_0)]$	$\boldsymbol{a}[x_0(1-\varphi_0)+(1+\varphi_0)]$	$\boldsymbol{2a(x_0+1)}$

Az x_0 értékei expliciten is meghatározhatók:

$$(x_0)_{1,2} = \frac{\varphi_0{}^2(1-\varphi^2)+(\varphi_0{}^2-\varphi^2)\pm 2\cdot\sqrt{\varphi_0{}^2(1-\varphi^2)(\varphi_0{}^2-\varphi^2)}}{(1-\varphi_0{}^2)\varphi^2}$$

Figyelembe véve hogy $Q = \frac{2\varphi_0}{1+\varphi_0{}^2}$, amiből $\varphi_0 = \frac{1-\sqrt{1-Q^2}}{Q}$, az x_0 az előre rögzített φ és Q függvényében a következőképpen adható meg:

$$(x_0)_{1,2} = \frac{1-\varphi^2-\sqrt{1-Q^2}\pm\sqrt{(1-\varphi^2)\left[\left(1-\sqrt{1-Q^2}\right)^2-\varphi^2Q^2\right]}}{\varphi^2\sqrt{1-Q^2}}$$

Ha az együttes gyakoriságok *pozitívak,* akkor adott φ és Q értékhez (ahol $\frac{2|\varphi|}{1+\varphi^2} \le |Q|$ és a két mutató előjele megegyezik) megadható[5] olyan eloszlás, amihez pont a megadott φ és Q értékek tartoznak:

$$f_{11} = a\frac{1-\varphi^2-\sqrt{1-Q^2}\pm\sqrt{(1-\varphi^2)\left[\left(1-\sqrt{1-Q^2}\right)^2-\varphi^2Q^2\right]}}{\varphi^2\sqrt{1-Q^2}}\cdot\frac{Q+1-\sqrt{1-Q^2}}{Q};$$

$$f_{12} = a\frac{1-\varphi^2-\sqrt{1-Q^2}\pm\sqrt{(1-\varphi^2)\left[\left(1-\sqrt{1-Q^2}\right)^2-\varphi^2Q^2\right]}}{\varphi^2\sqrt{1-Q^2}}\cdot\frac{Q-1+\sqrt{1-Q^2}}{Q};$$

$f_{21} = a\frac{Q-1+\sqrt{1-Q^2}}{Q}$ és $f_{22} = a\frac{Q+1-\sqrt{1-Q^2}}{Q}$, ahol a egy pozitív arányossági tényező.

II. Belátjuk, hogy van olyan együttes eloszlás bármely $0 < \varphi < 1$ esetén, amelyre a $Q = 1$ érték megvalósul. Ez a keresett eloszlások egyszerű megadásával történik, a *7. ábra* jobb oldali eloszlásai ilyenek. A bizonyításnak ehhez a részéhez nem tartozik, de a teljesség kedvéért a negatív φ értékekhez tartozó eloszlásokat is megadtuk (*7. ábra,* bal oldal). Tehát, ha az együttes gyakoriságok közül pontosan az egyik nulla, akkor a *7. ábrán* látható eloszlások olyanok, amikhez az előre adott φ érték tartozik, és a Q értéke 1 vagy –1.

7. ábra: *Adott φ-hez és Q = ± 1-hez tartozó eloszlások, ha a kapcsolat nem függvényszerű* (a > 0)

	Q = –1	
	B_1	B_2
A_1	0	$-a\varphi$
A_2	$-a\varphi$	$a(1+\varphi)$

	Q = 1	
	B_1	B_2
A_1	$a\varphi$	0
A_2	$a(1-\varphi)$	$a\varphi$

	Q = –1	
	B_1	B_2
A_1	$a(1+\varphi)$	$-a\varphi$
A_2	$-a\varphi$	0

	Q = 1	
	B_1	B_2
A_1	$a\varphi$	$a(1-\varphi)$
A_2	0	$a\varphi$

5 Ezzel az összes ilyen eloszlást nem adtuk meg.

Összegzésül elmondhatjuk, hogy az összes egyező előjelű φ és Q érték, amelyre $\frac{2|\varphi|}{1+\varphi^2} \leq |Q|$ teljesül, megvalósul valamely alkalmas eloszlás esetén.

A φ és a Q definíciói

A különböző asszociációs mutatók származtatásától azt várjuk el, hogy legyenek világosak a konstrukció alapvetései, és amennyiben ez lehetséges, a mutató definiálásából megállapítható legyen annak értelmezési módja. Sorra vesszük a φ és Q többféle megalapozását, és ezek segítségével próbálunk az értelmezéseikben adódó különbségekre következtetni. Először a mutatók azon bevezetési módjait elemezzük, melyek a mai felsőfokú statisztikaoktatásban használatosak: a Q esetén a koordinációs viszonyszámokból történő, a φ-nél pedig a χ^2 statisztikán alapuló levezetést. Mindkét módszer vizsgálatánál arra törekszünk, hogy a másik mutatót is az adott elvek alapján definiáljuk, tehát a φ-t is meghatározzuk koordinációs viszonyszámokkal, és a Q-nak is megadjuk egy olyan formáját, ami a függetlenség esetén feltételezhető gyakoriságokhoz való viszonyításon alapszik. Majd megadjuk a mutatóknak egy olyan származtatási módját, ami a Q mutató általánosításának tekinthető γ mutatónál használatos. Végül a φ és Q létrehozására eredetileg alkalmazott Yule-féle konstrukciós elveket vizsgáljuk meg. Mind a négy módszernél megállapítjuk, hogy megalapozható-e az adott módon mindkét mutató, és ha igen, akkor a konstrukció alapfogalmaival lehet-e természetes módon, a *statisztikai gyakorlat számára használható* interpretációját adni a mutatóknak. Amennyiben lehetséges ilyen értelmezéseket adni, meghatározzuk, hogy ezek az értelmezési módok hogyan teszik összehasonlíthatóvá a mutatókat.

A további vizsgálatok során az együttes eloszlás *8. ábrán* feltüntetett egyszerűsített jelölésmódját alkalmazzuk.

8. ábra: *Az együttes eloszlás és a függetlenség esetén fennálló eloszlás egyszerűsített jelölése*

	B_1	B_2	Összesen
A_1	a	b	$\boldsymbol{a+b}$
A_2	c	d	$\boldsymbol{c+d}$
Összesen	$\boldsymbol{d}$ $\boldsymbol{a+}$	$\boldsymbol{b+}$	$\boldsymbol{n}$

	Függetlenség esetén		
	B_1	B_2	Összesen
A_1	a*=	b*=	$\boldsymbol{a+b}$
A_2	c*=	d*=	$\boldsymbol{c+d}$
Összesen	$\boldsymbol{a+c}$	$\boldsymbol{b+d}$	$\boldsymbol{n}$

Definíció koordinációs viszonyszámokkal

A gazdasági felsőoktatás számára írt tankönyvekben a Yule-féle asszociációs együtthatót koordinációs viszonyszámok segítségével definiálják (Köves – Párniczky 1973: 302; Korpás 1996: 132; Kerékgyártó et al. 2001: 68; Sándorné et al. 2013: 139). A gondolatmenet a következő:

Az A és a B ismérvek akkor és csak akkor függetlenek, ha $\frac{a}{c} = \frac{b}{d}$ és $0 < a, b, c, d$.

Az egyenlőséget átrendezve:

$$ad - bc = 0$$

Érdemes megjegyezni, hogy függetlenség esetén egyik együttes gyakoriság (a, b, c, d) sem nulla, és mivel olyan eloszlásokat vizsgálunk, ahol a peremgyakoriságok pozitívak, az ekvivalencia valóban teljesül. Ha a függetlenség nem teljesül, az ismérvek között asszociációs kapcsolat van, $ad - bc \neq 0$, tehát $ad - bc$ normált értéke az asszociáció mértékének megfelelő mérőszáma lehet:

$$Q = \frac{ad - bc}{ad + bc}$$

A Q együttható koordinációs viszonyszámokkal történő bevezetése nehézségekbe ütközik, ha valamelyik együttes gyakoriság (a, b, c, d) nulla értékű. Függvényszerű kapcsolat esetén a kérdéses koordinációs viszonyszámok – nullával való osztás miatt – nem is értelmesek. Hosszas magyarázkodás után, határértékek vizsgálatával lehetne a koordinációs viszonyszámok segítségével precízen definiálni Q értékét. Amennyiben e módszer mellett döntünk, a Yule-féle asszociációs együttható a következő módon definiálható:

$$Q = \begin{cases} -1\,, ha\ a = 0\ vagy\ d = 0 \\ \dfrac{\frac{a}{c} - \frac{b}{d}}{\frac{a}{c} + \frac{b}{d}}, ha\ a, b, c, d \neq 0 \\ 1, ha\ b = 0\ vagy\ c = 0 \end{cases}$$

Ha a, b, c, $d \neq 0$, a Q koordinációs viszonyszámokkal a következő alakokban is írható:

$$Q = \frac{\frac{a}{c} - \frac{b}{d}}{\frac{a}{c} + \frac{b}{d}} = \frac{\frac{a}{b} - \frac{c}{d}}{\frac{a}{b} + \frac{c}{d}} = \frac{\frac{d}{c} - \frac{b}{a}}{\frac{d}{c} + \frac{b}{a}} = \frac{\frac{d}{b} - \frac{c}{a}}{\frac{d}{b} + \frac{c}{a}}$$

A törtek számlálója értelmezhető, két koordinációs viszonyszám különbségvételen alapuló összehasonlítását fejezi ki, ám a törtek nevezője koordinációs viszonyszámok összege, nincs gyakorlatban is használható értelmezése, ezért a Yule-féle asszociációs együtthatónak nincs a koordinációs viszonyszámokon alapuló gyakorlati értelmezése. A koordinációs viszonyszámokkal történő felírás olyankor lehet fontos, amikor hiányosak az együttes eloszlásról szóló ismereteink, csak a megfelelő koordinációs viszonyszámokat ismerjük. Például, ha a fent részletezett szintfelmérő feladatok teljesítéséről csak a következőket tudjuk: A *27. feladatot nem teljesítők körében 2,5385-ször többen nem teljesítették a 4. feladatot, mint ahányan teljesítették; a 27. feladatot teljesítők körében 5-ször annyian teljesítették a 4. feladatot, mint ahányan nem teljesítették.* Ezek alapján a Q értéke meghatározható: $Q = \frac{2{,}5385 - \frac{1}{5}}{2{,}5385 + \frac{1}{5}} = 0{,}8539.$

Bár a φ együtthatót nem szokás – és nem is érdemes – koordinációs viszonyszámok segítségével levezetni, a Q-val való összehasonlíthatóság érdekében megadjuk a koordinációs viszonyszámokkal felírt alakjait (a, b, c, $d \neq 0$):

$$\varphi = \frac{\frac{a}{c} - \frac{b}{d}}{\sqrt{\left(\frac{a}{c} + \frac{b}{d} + \frac{a}{d} + \frac{b}{c}\right)\left(\frac{a}{c} + 1\right)\left(\frac{b}{d} + 1\right)}} = \frac{\frac{a}{b} - \frac{c}{d}}{\sqrt{\left(\frac{a}{b} + \frac{c}{d} + \frac{a}{d} + \frac{c}{b}\right)\left(\frac{a}{b} + 1\right)\left(\frac{c}{d} + 1\right)}} =$$
$$= \frac{\frac{d}{c} - \frac{b}{a}}{\sqrt{\left(\frac{d}{c} + \frac{b}{a} + \frac{d}{a} + \frac{b}{c}\right)\left(\frac{d}{c} + 1\right)\left(\frac{b}{a} + 1\right)}} = \frac{\frac{d}{b} - \frac{c}{a}}{\sqrt{\left(\frac{d}{b} + \frac{c}{a} + \frac{d}{a} + \frac{c}{b}\right)\left(\frac{d}{b} + 1\right)\left(\frac{c}{a} + 1\right)}}$$

Ezekből a formulákból kiderül, hogy nem csak megalapozni nem érdemes a φ együtthatót koordinációs viszonyszámokkal, illetve az értelmezése sem lehetséges velük, de hiányos adatok esetén számításra is nagyon körülményesen használható. Négy megfelelő koordinációs viszonyszám ismeretében határozható meg az értéke, így technikai értelemben sem érdemes koordinációs viszonyszámokra alapozni a φ számítását. Elméleti szempontból lehet érdekes a φ ilyen megadása, ennek segítségével az abszolút értéke könnyen összehasonlítható a Q abszolút értékével:

$$|\varphi| = \frac{\left|\frac{a}{c} - \frac{b}{d}\right|}{\sqrt{\left(\frac{a}{c} + \frac{b}{d} + \frac{a}{d} + \frac{b}{c}\right)\left(\frac{a}{c} + 1\right)\left(\frac{b}{d} + 1\right)}} = \frac{\left|\frac{a}{c} - \frac{b}{d}\right|}{\sqrt{\left(\frac{a}{c} + \frac{b}{d} + \frac{a}{d} + \frac{b}{c}\right)\left(\frac{a}{c} \cdot \frac{b}{d} + \frac{a}{c} + \frac{b}{d} + 1\right)}} \leq \frac{\left|\frac{a}{c} - \frac{b}{d}\right|}{\sqrt{\left(\frac{a}{c} + \frac{b}{d}\right)\left(\frac{a}{c} + \frac{b}{d}\right)}} = |Q|$$

Definíció a függetlenség esetén fennálló gyakoriságoktól vett eltérésekkel

A φ mutató bevezetése *a* χ^2 statisztikán alapul, ezt szokás négyzetes kontingenciának is nevezni (Yule – Kendall 1964: 76), ami a függetlenség esetén feltételezhető gyakoriságok (a^*, b^*, c^*, d^*) és a tényleges gyakoriságok (a, b, c, d) eltérését méri. A Q mutatóra is megadható olyan alak, ami – *a* χ^2 statisztikától eltérő módon – a függetlenség esetén feltételezhető gyakoriságokhoz való viszonyítást használja.

A Q együtthatónak van egy figyelemre méltó tulajdonsága, miszerint tetszőleges α, β, γ, δ pozitív együtthatókra, amelyekre $\alpha\delta - \beta\gamma = 0$ teljesül,

$$Q = \frac{ad - bc}{ad + bc} = \frac{\alpha a \delta d - \beta b \gamma c}{\alpha a \delta d + \beta b \gamma c} = \frac{\frac{a}{\alpha}\frac{d}{\delta} - \frac{b}{\beta}\frac{c}{\gamma}}{\frac{a}{\alpha}\frac{d}{\delta} + \frac{b}{\beta}\frac{c}{\gamma}}$$

Az összefüggés egyszerűen belátható, mert ha $\alpha\delta - \beta\gamma = 0$, akkor $\alpha\delta = \beta\gamma = \varepsilon > 0$. Egyrészt bővítve, másrészt egyszerűsítve a törtet ε-nal, a fenti egyenlőséghez jutunk.

Az $\alpha = a^*$, $\beta = b^*$, $\gamma = c^*$, $\delta = d^*$ választással kapjuk:

$$Q = \frac{\frac{a}{a^*} \cdot \frac{d}{d^*} - \frac{b}{b^*} \cdot \frac{c}{c^*}}{\frac{a}{a^*} \cdot \frac{d}{d^*} + \frac{b}{b^*} \cdot \frac{c}{c^*}}$$

Az $\frac{a}{a^*}, \frac{d}{d^*}, \frac{b}{b^*}, \frac{c}{c^*}$ törtek hányados képzésen alapuló összehasonlítást fejeznek ki, azt mutatják meg, hogy a tényleges gyakoriságok hányszorosai a függetlenség esetén fennálló gyakoriságoknak; viszont a belőlük képzett mutató pusztán ezekre az arányokra hivatkozva a statisztikai gyakorlatban nem értelmezhető. Ez a fajta felírás csak a Q kiszámítását teszi lehetővé az adott arányok ismeretében.

A φ esetén más a helyzet, mert az eredeti meghatározásban szereplő $\varphi^2 = \frac{\chi^2}{n}$ formula (Yule – Kendall 1964: 76), a φ-nek előjelet tulajdonítva, többféleképpen is átfogalmazható:

(1)
$$\varphi = sign(ad - bc) \cdot \sqrt{\frac{a^*\left(\frac{a-a^*}{a^*}\right)^2 + b^*\left(\frac{b-b^*}{b^*}\right)^2 + c^*\left(\frac{c-c^*}{c^*}\right)^2 + d^*\left(\frac{d-d^*}{d^*}\right)^2}{a^* + b^* + c^* + d^*}}$$

(2)
$$\varphi = sign(ad - bc) \cdot \sqrt[4]{\left(\frac{a-a^*}{a^*}\right)\left(\frac{b-b^*}{b^*}\right)\left(\frac{c-c^*}{c^*}\right)\left(\frac{d-d^*}{d^*}\right)}$$

Mindkét formula a gyakoriságok és a függetlenség esetén fennálló gyakoriságok relatív eltéréseinek az átlagát fejezi ki. Az (1) képlet súlyozott négyzetes, a (2) súlyozatlan mértani[6] átlag. Gyakorlati szempontból értelmezhető mindkét felírás, a φ mutató többek között a gyakoriságok átlagos relatív eltérését mutatja a függetlenség esetén fennálló gyakoriságoktól. Tekintsük a következő példát! *A szintfelmérő 4. és 27. feladata esetén a hallgatók teljesítményéről az 2. táblázatban foglaltakat ismerjük*.

A *2. táblázat* adataiból:

$$Q = \frac{1{,}0972 \cdot 2{,}4074 - 0{,}2549 \cdot 0{,}8164}{1{,}0972 \cdot 2{,}4074 + 0{,}2549 \cdot 0{,}8164} = 0{,}8539 \text{ és}^7\ \varphi = 1 \cdot \sqrt[4]{0{,}0972 \cdot 0{,}7451 \cdot 0{,}1836 \cdot 1{,}4071} = 0{,}3698.$$

6 A negyedik gyök alatt a törtek abszolút értékét kellene szerepeltetni, mert ezeknek vesszük a mértani átlagát, egyébként a tö rtek különböző előjelűek, és mint ilyeneknek, nem is értelmes a mértani átlaguk. De mivel a törtek közül kettő-kettőnek megegyezik az előjele, az abszolútérték-jel elhagyható.

7 $sign(ad–bc) = sign(1{,}0972 \cdot 2{,}4074–0{,}2549 \cdot 0{,}8164) = 1$.

A φ jelen felépítésből adódó értelmezése: *A szintfelmérő 4. és 27. feladatának teljesítését vizsgálva, a hallgatók kombinatív osztályozással nyert, teljesítmény szerinti csoportjaiban a hallgatók száma átlagosan 36,98%-kal tér el a függetlenség esetén feltételezhető létszámtól.*

2. táblázat: *A hallgatók számának relatív eltérése (%) a függetlenség esetén feltételezhető gyakoriságoktól a 4. és 27. feladat teljesítése szerint*

		27. feladat	
		Nem teljesítette	**Teljesítette**
4. feladat	Nem teljesítette	+ 9,72	–74,51
	Teljesítette	–18,36	+140,74

Forrás: saját szerkesztés (2014)

Definíció, ami az ellentétes irányú asszociáción alapul

Az asszociáció vizsgálata során az ismérvek kapcsolatának három esetét különböztetjük meg: (1) függetlenség esetén az asszociáció teljes hiányát, (2) pozitív asszociációt (vagy röviden asszociációt) és (3) negatív asszociációt (disszociációt). Ezekben az esetekben, ha f jelöli a 2×2-es tábla valamelyik együttes gyakoriságát és f^* a hozzá tartozó függetlenség esetén fennálló gyakoriságot,

$f = f^*$, ha függetlenség,

$f > f^*$, ha pozitív asszociáció és

$f < f^*$, ha negatív asszociáció tapasztalható a megfelelő ismérvváltozatoknál.

Tehát az $f - f^*$ különbség előjele megadja az asszociáció irányát.

A *8. ábra* jelöléseit használva $a - a^* = a - \frac{(a+b)(a+c)}{n} = \frac{ad-bc}{n} = D$, innen

$$\begin{array}{ll} a - a^* = D & \quad b - b^* = -D \\ c - c^* = -D & \quad d - d^* = D\ . \end{array}$$

Ezek alapján látható, hogy a 2×2-es táblák két átlójában ellentétes irányú asszociáció tapasztalható. A következő gondolatmenet az azonos előjelű asszociációs kapcsolatban lévő párok leszámlálásán alapul *(9. ábra).* A *9. ábra* bal oldali eloszlásához (önkényesen) 1-et rendelünk az asszociáció mértékéül, az ellentétes kapcsolatot mutató jobb oldalon látható eloszláshoz –1-et.

9. ábra: *Azonos előjelű asszociációs kapcsolatban lévő párok eloszlásai*

	Az asszociáció mértéke 1		
	B_1	B_2	**Összesen**
A_1	1	0	***1***
A_2	0	1	***1***
Összesen	***1***	***1***	***2***

	Az asszociáció mértéke –1		
	B_1	B_2	**Összesen**
A_1	0	1	***1***
A_2	1	0	***1***
Összesen	***1***	***1***	***2***

A *8. ábrán* látható eredeti eloszlás alapján $a \cdot d$ olyan pár van, melyeknek az asszociációs mértéke 1, és $b \cdot c$ olyan, aminek –1.

A Yule-féle asszociációs együttható a következő formákban írható:

(1) $$Q = \frac{ad}{ad + bc} - \frac{bc}{ad + bc}$$

(2) $$Q = 1 \cdot \frac{ad}{ad + bc} + (-1) \cdot \frac{bc}{ad + bc}$$

Az (1) felírás megoszlási viszonyszámok különbségvételen alapuló összehasonlításaként értelmezhető, megmutatja, hogy az azonos előjelű asszociációs mértékkel rendelkező párok[8] körében mennyivel nagyobb azoknak a pároknak az aránya, amiknek az asszociációs mértéke 1, mint azoknak, amiknek –1. A (2) alak súlyozott számtani átlagként értelmezhető, megadja, hogy mennyi az asszociáció mértékének átlaga az asszociációs kapcsolatban lévő párok körében.

8 Az itt definiált párok a Goodman-féle γ konstrukciójánál alkalmazott konkordáns és diszkonkordáns párok analogonjai (Kotz 2006: 2868–2869).

A φ felírásához először bevezetünk két mennyiséget:

$$\varphi_1 = \frac{a}{a+b} - \frac{c}{c+d} = \frac{ad}{(a+b)(c+d)} - \frac{bc}{(a+b)(c+d)} = 1 \cdot \frac{ad}{(a+b)(c+d)} + (-1) \cdot \frac{bc}{(a+b)(c+d)}$$

$$\varphi_2 = \frac{a}{a+c} - \frac{b}{b+d} = \frac{ad}{(a+c)(b+d)} - \frac{bc}{(a+c)(b+d)} = 1 \cdot \frac{ad}{(a+c)(b+d)} + (-1) \cdot \frac{bc}{(a+c)(b+d)}$$

A φ_1 és φ_2 önállóan is értelmezhető és használatos[9] mutatók. A φ_1 lehetséges értelmezései: (1) Megadja, hogy mennyivel nagyobb az A_1 ismérvváltozattal rendelkezők körében a B_1 ismérvváltozathoz tartozók aránya, mint az A_2 ismérvváltozattal rendelkezők körében. (2) Megmutatja, hogy az olyan párok körében, amelyeknek egyike az A_1, a másika az A_2 ismérvváltozattal rendelkezik, mennyivel nagyobb azoknak a pároknak az aránya, amiknek az asszociációs mértéke 1, mint azoknak, amiknek –1. (3) Kifejezi, hogy mennyi az asszociáció mértékének átlaga az olyan párok körében, amelyeknek az egyik tagja az A_1, a másik tagja az A_2 ismérvváltozattal rendelkezik. Mivel φ_1 és φ_2 előjele megegyezik, értelmes a következő formula (a négyzetgyök alatti kifejezés nemnegatív):

$$\varphi = sign(ad - bc) \cdot \sqrt{\varphi_1 \cdot \varphi_2}$$

Megjegyezzük, hogy $sign(ad - bc) = sign(\varphi_1) = sign(\varphi_2)$.

A φ mutató a φ_1 és φ_2 mértani átlagaként[10] értelmezhető: megadja, hogy az olyan párok esetén, amelyek tagjai különböző ismérvváltozattal rendelkeznek, mennyi az asszociáció mértékének átlaga, illetve átlagosan mennyivel nagyobb azoknak a pároknak az aránya, amiknek az asszociációs mértéke 1, mint azoknak, amiknek –1.

9 A kismintás függetlenségvizsgálatoknál alkalmazott egzakt próbáknál használják asszociációs mértéknek.

10 A mértani átlagot nemnegatív átlagolandó értékekre szokás értelmezni, de a súlyozatlan mértani átlag kiterjeszthető nempozitív értékekre is, ebben az esetben az átlag értéke nempozitív. A súlyozatlan mértani átlag olyan esetben nem használható, ha az átlagolandó értékek között van negatív és pozitív is.

Definíció a teljes asszociáció fogalma alapján

A teljes asszociációval történő megalapozás alapgondolata az, hogy az asszociációs mérték függetlenség esetén 0, teljes asszociáció esetén az 1 vagy a –1 értéket vegye fel. A köztes esetekben e szélső értékek közti eredményt adjon, a 0-hoz közeli értékek a függetlenség esetén fennálló együttes eloszlásokhoz közeli, a ±1-hez közeli értékek a teljes asszociációt megvalósító eloszlásokhoz közeli esetekben valósuljanak meg.

A függetlenség definíciója – az ekvivalens megfogalmazásoktól eltekintve – a szakirodalomban egységes, ezzel szemben a teljes asszociációt többféleképpen is lehet definiálni.

A diszjunktív teljes asszociáció definíciója: *Két ismérv esetén diszjunktív teljes asszociációról beszélünk abban az esetben, ha az egyik ismérv legalább egyik változata szükségképpen együtt jár a másik ismérv valamelyik változatával* (Yule – Kendall 1964: 47). Egy egyszerű példával illusztrálva: A hallgatók egy csoportjáról tudjuk, hogy a csoportban minden fiú dohányzik. Ebben az esetben a *nem és a dohányzási szokás* esetén diszjunktív teljes asszociációt tapasztalunk, mert a *nem* egyik ismérvváltozata – az, hogy fiú – szükségképpen maga után vonja a másik ismérv, a *dohányzási szokás* egyik változatát – azt, hogy dohányzik. Ez nem jelenti azt, hogy minden dohányzó szükségképpen fiú is. A diszjunktív teljes asszociációt legegyszerűbben a 0 gyakoriságok elhelyezkedésével lehet leírni, a *10. ábra* ezt szemlélteti. Az ábrán a színezett mezők a pozitív gyakoriságok helyét jelölik.

10. ábra: *A diszjunktív teljes asszociáció esetei*

	B_1	B_2
A_1	0	
A_2		

	B_1	B_2
A_1		
A_2		0

	B_1	B_2
A_1	0	
A_2		0

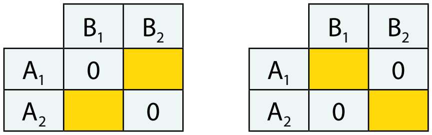
	B_1	B_2
A_1		0
A_2		

	B_1	B_2
A_1		
A_2	0	

	B_1	B_2
A_1		0
A_2	0	

A konjunktív teljes asszociáció definíciója: *Két ismérv esetén konjunktív teljes asszociációról beszélünk abban az esetben, ha az egyik ismérv mindegyik változata szükségképpen együtt jár a másik ismérv valamelyik változatával.* A teljes asszociációnak ezt a fogalmát szokták abszolút asszociációnak is nevezni (Kendall – Stuart 1961: 540), illetve megállapíthatjuk, hogy ekvivalens a *függvényszerű kapcsolat* fogalmával. A konjunktív teljes asszociáció lehetséges elrendezéseit mutatja a *11. ábra,* a színezett mezők itt is pozitív gyakoriságokat szimbolizálnak.

11. ábra: *A konjunktív teljes asszociáció (függvényszerű kapcsolat) esetei*



A teljes asszociációnak tehát kétféle definíciójáról, kétféle értelmezéséről beszélünk. Olyan mutatókat keresünk, amelyek ezen meghatározásokon alapulnak, az abszolút értékük 1 a teljes asszociáció, 0 a függetlenség esetén. A Yule-féle asszociációs együttható és a φ ilyen mutatók.

A Yule-féle Q együttható a diszjunktív teljes asszociáción alapuló mutató:

$Q = 0$ akkor és csak akkor áll fenn, ha az ismérvek *függetlenek;*

$|Q| = 1$ akkor és csak akkor teljesül, ha *diszjunktív teljes asszociációt* tapasztalunk az ismérveknél, tehát az ismérvek legalább egyik változata minden esetben együtt jár a másik ismérv valamelyik változatával;

$|Q|$ 1-hez közeli értékei azt fejezik ki, hogy az ismérvek legalább egyik ismérvváltozata szinte minden esetben együtt jár a másik ismérv valamelyik változatával, az előjele segítségével pedig meghatározhatjuk, hogy melyek ezek az ismérvváltozatok.

A φ mutató a konjunktív teljes asszociáción alapul:

$\varphi = 0$ akkor és csak akkor áll fenn, ha az ismérvek *függetlenek;*

$|\varphi| = 1$ akkor és csak akkor teljesül, ha *konjunktív teljes asszociációt,* más néven *függvényszerű kap-*

csolatot tapasztalunk az ismérveknél, tehát az ismérvek mindkét változata minden esetben együtt jár a másik ismérv valamelyik változatával;

$|\varphi|$ 1-hez közeli értékei azt fejezik ki, hogy az ismérvek mindkét ismérvváltozata szinte minden esetben együtt jár a másik ismérv valamelyik változatával, az előjelének figyelembevételével következtethetünk ezekre az ismérvváltozatokra.

A teljes asszociáció két eltérő fogalmával a φ és Q definiálható, a teljes asszociáció különböző értelmezéseit figyelembe véve a gyakorlatban alkalmazott vizsgálatok során is jól interpretálhatók. Mivel az adott ismérvek kapcsolatát más értelemben mérik, de jelen megalapozásból következően összeegyeztethető módon, ezért adott esetben érdemes mindkét mutatót meghatározni, egymást kiegészítő értelmezéseiket adni.

A két mutató javasolt értelmezését a bevezető példa adatain demonstráljuk. Ahogy azt már az előzőekben meghatároztuk, a szintfelmérő dolgozat 4. és 27. feladatának teljesítése közti kapcsolatot vizsgálva $Q = 0{,}8539$ és $\varphi = 0{,}3698$. A két mutató értéke jelentősen eltér egymástól, és ez az eltérés φ nullához közeli értékei estén tipikusnak mondható, ha valamelyik ismérvváltozat a teljes asszociációhoz közeli eloszlást mutat. A φ alacsony értéke arra utal, hogy nincs mindkét ismérvváltozat szerint erős kapcsolat, a Q magas értéke pedig arra, hogy legalább az egyik változat szerint erős kapcsolat tapasztalható. E két információt együttesen úgy értelmezhetjük, hogy – mindkét ismérv esetén – pontosan egy ismérvváltozatnál tapasztalható erős kapcsolat.

A φ mutató alapján elmondható, hogy a hatványozási azonosság (4. feladat) és a gyöktelenítés (27. feladat) teljesítése között közepesnél gyengébb sztochasztikus kapcsolat tapasztalható, az együttes gyakoriságok átlagosan 36,98%-kal térnek el a függetlenség esetén várható együttes gyakoriságoktól.

A Yule-féle asszociációs együttható (Q) szerint kijelenthető, hogy a feladatok teljesítésének csak az egyik kimenetele mutat erős sztochasztikus kapcsolatot a másik feladat teljesítésével. Az eloszlás alapján (1. táblázat): akik nem tudták a hatványozási azonosságot, azoknak 96,88%-a nem tudta a gyöktelenítést; és akik tudtak gyökteleníteni, azoknak a 83,33%-a tudta a hatványozási azonosságot.

Összegzés

Az elemzések során kiderült, hogy a Yule-mutató abszolút értéke legalább annyi, mint $\frac{2|\varphi|}{1+\varphi^2}$, az egyenlőség csak szimmetrikus 2×2-es táblák, függetlenség és függvényszerű kapcsolat esetén teljesül; tehát a Yule-mutató minden más esetben erősebb asszociációs kapcsolatot jelez, mint a φ. A két mutató definíciói közül a teljes asszociáció fogalmán alapuló meghatározás alkalmazható mindkét mutatóra oly módon, hogy az értelmezésük alapján összehasonlíthatóak legyenek, a többi vizsgált definíció erre alkalmatlannak bizonyult. A φ és a Yule-mutató szükségképpen fennálló eltérésére tehát a magyarázatot az adja, hogy míg a vizsgálatba vont változók függetlenségét a két mutató azonos módon veszi figyelembe, a teljes asszociációt másként. Az összehasonlítást lehetővé tevő definíció szerint mindkét mutató konstrukciójának alapgondolata, hogy az asszociációs kapcsolat mértékét a teljes asszociáció és a függetlenség között mérje. A Yule-féle konstrukciós alapfeltevések vizsgálata során nyilvánvalóvá vált, hogy a teljes asszociáció definíciója eltér a két mutató esetén: A Yule-féle meghatározás szerint teljes asszociációról akkor beszélünk, ha legalább egy ismérvváltozat szükségképpen együtt jár a másik ismérv valamelyik változatával, és a teljes asszociáció ezen meghatározásának tesz eleget a Yule-féle asszociációs együttható (Yule 1900, 1912; Yule – Kendall 1964). A φ meghatározásánál a teljes asszociáció egy másik meghatározása játszik szerepet, ebben az esetben teljes asszociációról akkor beszélünk, ha *mindegyik* ismérvváltozat szükségképpen együtt jár a másik ismérv valamelyik változatával. A Yule-mutató és a φ a teljes asszociáció mértékét más-más értelemben méri, amit e mutatók értelmezésénél és használatánál figyelembe kell venni.

A felsőfokú gazdasági statisztikaoktatásban a 2×2-es táblák asszociációs vizsgálatára a Yule-féle asszociációs együtthatót javasolják, mert egyszerű kiszámolni. A Csuprov-mutató esetén is ismert egyszerű formula, ezért a számítási bonyolultság nem indokolja, hogy vele szemben a Yule-mutatót előnyben részesítsük. Az asszociációs vizsgálatoknál nem is lehet elfogadható indok a választott mérőszám egyszerűsége, mert két ismérv kapcsolatáról többféle értelemben beszélhetünk, a Csuprov- és a Yule-mutató más-más információt szolgáltat az együttes eloszlásról. Érdemes megfontolni, hogy a mindennapi statisztikaoktatásban a 2×2-es táblák vizsgálata során mindkét mutatót tanítsuk és számoltassuk a hallgatókkal. Amennyiben a két mutató közel azonos abszolút értékű (ez csak gyenge és erős kapcsolat esetén fordulhat elő), a két asszociációs mérték egyszerre értelmezhető; ha a két mutató abszolút értéke jelentősen eltér egymástól, akkor a két mutató logikailag egymásra épülő elemzésével készíthetünk értelmezést, és az együttes eloszlás alapján rávilágíthatunk eltérésük okára.

Kitekintés

A Kotz-féle Statisztikai tudományok enciklopédiája (Kotz 2006) a Yule-féle Q-t, a kolligációs együtthatót, a φ-t, a Csuprov- és Cramer-mutatókat a „régi asszociációs mértékek" közé sorolja, velük szemben a Goodman–Kruskal-féle λ-t helyezi előtérbe. Az asszociációs mértékeknek több családja ismert: a χ^2 statisztikán, feltételes valószínűségeken és információelméleten alapuló mérőszámok. Ezen mutatók konstrukciós elve, használati köre és értelmezési módja is eltér egymástól, tehát semmiképp sem gondolhatjuk, hogy a λ az egyedüli modern asszociációs mérték (ráadásul közismert az a hátrányos tulajdonsága, hogy nem csak függetlenség esetén ad 0 értéket). A pszichológiai, pedagógiai és biológiai kutatások során gyakran szembesülünk azzal, hogy a vizsgálatba vont változókról kevés megfigyeléssel rendelkezünk. Ilyenkor nem alkalmazhatók a nagymintás, közelítő eloszlásokon alapuló módszerek, gyakran csak a kismintás, a pontos eloszlást használó (egzakt) következtetési módszereket alkalmazhatjuk. Az egzakt próbák olykor igen számításigényesek, papíron ceruzával nem hajthatók végre, de a számítógépek teljesítményének fejlődésével elhárult az akadály a kismintás függetlenségvizsgálatra konstruálható egzakt próbák végrehajtása elől. A napjainkban kidolgozott egzakt próbák (Bereger 1996; Ruxton – Neuhauser 2010) többsége valamilyen asszociációs mértéken alapul, ezért alapvető fontosságú e mértékek alapos ismerete, illetve annak tisztázása, hogy a „régi" asszociációs mutatók az egzakt próbák során miként alkalmazhatók. További kutatások gazdag forrása lehet tehát annak a tisztázása, hogy kismintás függetlenségvizsgálatok esetében adott tudományos kérdés eldöntésére milyen asszociációs mérték és milyen egzakt próba használata indokolt.

Irodalomjegyzék

Berger, R. (1996): More powerful tests from confidence interval p values. *American Statistician,* 50: 314–318.

Ekström, J. (2008): The Phi-coefficient, the Tetrachoric Correlation Coefficient, and the Pearson-Yule Debate, http://statistics.ucla.edu/preprints/uclastat-preprint-2008:40 (letöltve: 2014. február 23.).

Ekström, J. (2009): An Empirical Polychoric Correlation Coefficient, http://statistics.ucla.edu/preprints/uclastat-preprint-2009:22 (letöltve: 2014. február 23.).

Goodman, L. A. – Kruskal, W. H. (1979): *Measures of Association for Cross Classifications.* New York: Springer-Verlag.

Kendall, M. G. – Stuart, A. (1961): *The Advanced Theory of Statistics.* Vol. 2. New York: Hafner Publishing Company.

Kerékgyártó Gy.-né – Mundruczó Gy. – Sugár A. (2001): *Statisztikai módszerek és alkalmazásuk a gazdasági, üzleti elemzésben.* Budapest: Aula Kiadó.

Korpás A.-né (1996): *Általános statisztika I.* Budapest: Nemzeti Tankönyvkiadó Rt.

Kotz, S. (ed.) (2006): *Encyclopedia of Statistical Sciences.* New Jersey: John Wiley & Sons, Inc., Hoboken.

Köves P. – Párniczky G. (1973): *Általános statisztika.* Budapest: Közgazdasági és Jogi Könyvkiadó.

Ruxton, G. D. – Neuhauser, M. (2010): Good practice in testing for an association in contingency tables. *Behavioral Ecology and Sociobiology,* 64: 1505–1513.

Sándorné K. É. – Csesznák A. – Ország G.-né (2013): *Statisztika I.* Budapest: Nemzedékek Tudása Tankönyvkiadó.

Tan, P.-N. – Kumar, V. – Srivastava, J. (2004): Selecting the right objective measure for association analysis. *Information Systems,* 29: 293–313.

Yule, G. U. (1900): On the Association of attributes in statistics: With illustrations from the material of the childhood society, &c. Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a *Mathematical or Physical Character,* 194: 257–319.

Yule, G. U. (1912): On the methods of measuring association between two attributes. *Journal of the Royal Statistical Society,* 75(6): 579–652.

Yule, G. U. – Kendall, M. G. (1964): *Bevezetés a statisztika elméletébe.* Budapest: Közgazdasági és Jogi Könyvkiadó.