

MTA Számítástechnikai és Automatizálási Kutató Intézet

Egymással kölcsönhatásban lévő betegségek erősségének új mérési lehetőségei

Börzsönyi László

Az orvos-, bio- és agrotudományokban gyakran vetődik fel a kérdés, hogy egy adott populáció esetében, két egymással összefüggésben lévő betegség kapcsolata milyen. Még pontosabban megfogalmazva a kérdést, a betegségek lefolyásának egyes stádiumait tekintve, hogyan jellemezhetők és mérhetők a kölcsönösen ható betegségek erősségei?

Milyen legyen az ellenanyagkezelés folyamata olyan betegségek esetén, melyek pillanatnyi ismereteink szerint nem szüntethetők meg, de alkalmas hatóanyagkezeléssel a káros hatások viszonylag jól kompenzálhatók?

Karakterizálják a  $\xi$  és  $\eta$  valószínűségi változók a vizsgálandó két betegséget. Feltéve, hogy a megfelelő várható értékek léteznek, a betegségek közötti kapcsolat erősségét a  $\xi$  és  $\eta$  változók kovariációja vagy még világosabban az

$$R(\xi, \eta) = \frac{M[(\xi - M(\xi))(\eta - M(\eta))]}{D(\xi)D(\eta)} \quad / 1 /$$

korrelációs koefficiens fejezi ki.

Ha a  $\chi^2$ -próba 1 szabadságfoknál szignifikáns, akkor használható a Pearson-féle

$$V = \frac{ad - bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}} \quad / 2 /$$

négymezős korrelációs együttható, vagy a

$$r_{\text{tet}} = \cos \pi \frac{\sqrt{bc}}{\sqrt{ad} + \sqrt{bc}} \quad / 3 /$$

tetrakorrelációs együttható, ami /2/-t jól közelíti. /2/-ben és /3/-ban az a, b, c, d változók a következő 2x2-es tábla elemeit jelentik,

		Y		
		igen	nem	
X	igen	a	b	/ 4 /
	nem	c	d	

ahol most X és Y a betegségek alternatív jellemzői.

Az /1/, /2/, /3/ formulákkal kifejezett fogalmak adnak egy bizonyos információt a betegségek közötti kapcsolat szorosságáról, azonban van egy hátrányuk is. Nevezetesen az, hogy az /1/ függvény  $\xi$ -ben és  $\eta$ -ban, a /2/ és /3/ függvények pedig b-ben és c-ben szimmetrikusak. Tehát arra a kérdésre, hogy a vizsgált kapcsolatban melyik betegség dominánsabb és milyen mértékben, az előző fogalmak nem adnak felvilágosítást. Továbbá az  $R(\xi, \eta) = 0$  (vagy  $\approx 0$ ) esetben azt mondjuk, hogy a  $\xi$  és  $\eta$  valószínűségi változók korrelálatlanok. Ebből nem következik az, hogy a betegségek függetlenek. Csupán azt állíthatjuk, hogy a betegségek között lineáris kapcsolat nem állhat fent, de ettől függetlenül közöttük még nagyon szoros nemlineáris kapcsolat létezhet.

A fölvetett probléma megválaszolásához egy asszimmetrikus 2x2-es tábla mértékre van szükségünk. Ilyen mérték kidolgozásával először G. Deuchler próbálkozott és az (1), (2) munkákban a következő

$$E = \frac{ad - bc}{(a+b)(c+d)} \quad / 5 /$$

mérőszámot közölte.

Youden 1950-ben, Ewards 1957-ben egymástól és Deuchlertől függetlenül szintén publikálták az /5/ mértéket. Hellmich 1969-ben a következő

$$K = \frac{ad - bc}{\sqrt{(a^2 + b^2)(c^2 + d^2)}}$$

/6/

asszimmetrikus 2x2-es tábla mértéket adta meg.

A Hellmich-féle K-mérték a következő módon általánosítható. Tekintsük a /4/ 2x2-es táblát. Az Y előfordulásának valószínűsége X-nél  $\frac{a}{a+b}$ . A nem-Y alternatíva előfordulásának valószínűsége X-nél:  $\frac{b}{a+b}$ . Teljesen analóg Y és nem-Y előfordulásának valószínűsége nem-X-nél  $\frac{c}{c+d}$ , illetve  $\frac{d}{c+d}$ . Értelmezzük a kapott mennyiségeket úgy, mint síkbeli vektorok első, illetőleg második komponenseit. Tehát a vektorok:

$$\bar{v}_1 = \left( \frac{a}{a+b}, \frac{b}{a+b} \right), \quad \bar{v}_2 = \left( \frac{c}{c+d}, \frac{d}{c+d} \right).$$

Általánosított K-mérték alatt a  $\bar{v}_1$  és  $\bar{v}_2$  síkbeli szabad vektorok által bezárt szög szinuszát értjük:

$$K = \sin(\bar{v}_1, \bar{v}_2) = \sin \varphi.$$

A  $\sin \varphi$  értékét a vektorok külső szorzására vonatkozó

$$|\bar{v}_1 \times \bar{v}_2| = |\bar{v}_1| \cdot |\bar{v}_2| \cdot \sin(\bar{v}_1, \bar{v}_2)$$

identitásból számíthatjuk. Szómszerűleg K-et kapunk, ami mindig nemnegatív, így K-ra a következő feltételeket tesszük:

$$\begin{array}{lll} K > 0 & \text{ha} & ad > bc \\ K < 0 & \text{ha} & ad < bc. \end{array}$$

K asszimmetrikus volta miatt b és c fölcserélésével két értéket kapunk:

$$K_1 = \frac{ad - bc}{\sqrt{(a^2 + b^2)(c^2 + d^2)}}, \quad K_2 = \frac{ad - bc}{\sqrt{(a^2 + c^2)(b^2 + d^2)}}.$$

A  $K_1$  és  $K_2$  értékek tulajdonságai:

$$\begin{aligned} -1 &\leq K_1, K_2 \leq +1 \\ &> \\ K_1, K_2 &\leq 0 \quad \text{egyidejűleg.} \end{aligned}$$

A  $K_1$  és  $K_2$  értékek határesetei a következők:

1.  $K_1 = -1, K_2 = -1$  legerősebb kölcsönös elkülönülés.
2.  $K_1 = 0, K_2 = 0$  teljes kölcsönös kapcsolatnélküliség.
3.  $K_1 = +1, K_2 = +1$  legerősebb kölcsönös összetartozás.

A Hellmich-féle K-mérték alkalmazható az összefüggésvizsgálat mindazon módszereiben, ahol a kapcsolat erősségét négymezős korrelációs együtthatóval mérjük. Az asszimmetricitás miatt ezen mérőszámok gazdagabb információt adnak az összetartozás módjáról, komplexebb formában jellemzik az adott kapcsolatot. Tekintsük a következő két problémát.

A.) Klasszikus irodalmi adatok alapján ismeretes, hogy a kukorica két betegsége a szárcorhadás és a csőpenész elhanyagolhatóan kicsiny korrelációt mutatnak. Azonban egyrészt az egyre több jó tulajdonsággal rendelkező nemesített fajták rezisztenciája a nemesítés során megváltozhatott, másrészt a termést igen hatásosan befolyásoló vegyszeres kezelések a talaj kémiai összetételét is átalakíthatták. Továbbá a természeti tapasztalatok is azt mutatják, hogy az egyik betegség bizonyos stádiumaiban a másik betegség is erőteljesebben jelentkezik. Célszerű tehát annak a kérdésnek a vizsgálata, hogy változott-e a betegségek kapcsolata az említett körülmények között és hogyan.

212 különböző hibrid kukorica mindegyikét 10 növényrel vetették el az alábbi körülmények között: \*

---

\* A szántóföldi kísérletet Dr. Manninger István kandidátus végezte, melynek adatait volt szives használatra bocsátani.

I. A csöveket mesterségesen fertőzték, térben nem szigetelték és a talajt NPK keverékkel kezelték. A kikelt növényeken összesen 1904 db csövet vizsgáltak meg.

II. A csöveket mesterségesen nem fertőzték, térben nem szigetelték, a talajt NPK keverékkel kezelték és összesen 1980 db csövet vizsgáltak meg.

III. A csöveket mesterségesen nem fertőzték, de a beteg csöveket térben szigetelték és a talajt NPK keverékkel kezelték. A megvizsgált csövek száma 1896 db volt.

IV. A csöveket mesterségesen nem fertőzték, a beteg csöveket viszont térben szigetelték. A talajt nitrogénnel (N) kezelték és 2109 db csövet vizsgáltak meg.

V. A csöveket mesterségesen nem fertőzték, a beteg csöveket térben szigetelték, a talajt pedig évtizedek óta nem kezelték. Ebben az esetben 1865 db csövet vizsgáltak meg.

Bonitáljuk az I., ..., V. esetekben a megvizsgált csöveket 1-től 9-ig úgy, hogy 1-es értéket kapjon a teljesen egészséges cső, 5-ös értéket a félig csőpenészes cső és 9-es bonitálási értéket a teljesen beteg cső. Hasonlóan rendre 1, 5 és 9 bonitálási értékeket kapjanak az egészséges, félig megdől és teljesen megdől, földön fekvő száraz. A további 1 - 9 közötti bonitálási értékek az előzőekhez viszonyított megbetegedett csőfelületeket, illetőleg szármegdőléseket fejezik ki. A bonitálás mindegyik esetben egy osztályozást ad. Az osztályok diszjunktak és számuk mindkét betegség esetén 9. Ha a betegségek osztályait eseményeknek fogjuk föl, akkor tulajdonképpen két teljes eseményrendszer függetlenségvizsgálatáról van szó.

Rendezzük a mért adatokat az I., ..., V. esetek mindegyikében egy  $9 \times 9$ -es kontingencia-táblába. A több osztályos kvalitatív változók összefüggésvizsgálatát a szokásos módon végezzük el, majd  $\chi^2$ -próbával döntünk a betegségek függetlenségéről. Határozzuk meg továbbá a  $K_1$  és  $K_2$  értékek közös maximumát az alábbi két feltétel egyidejű teljesülésével:

$$\max_i (K_1^i + K_2^i) \quad \text{és} \quad \max_i |K_1^i - K_2^i|$$

ahol  $i = 1, \dots, 81$ .

A vizsgálat összefoglaló eredményeit az 1. táblázat tartalmazza. Az eredmények azt mutatják, hogy NPK talajkezelés esetén a szárkorhadás és csőpenész egymással szignifikánsan összefüggnek. Ha a csövek térben szigeteltek és nitrogénnel (N) kezeljük a talajt vagy egyáltalán nem kezeljük a talajt, akkor a betegségek között nincs szignifikáns kapcsolat. A kezeletlen talaj esete igen jól alátámasztja a korábbi irodalmi adatokat. Az NPK talajkezelés viszont úgy tűnik, hogy befolyásolja a betegségek kapcsolatát, hiszen térben nem szigetelt esetben  $P = 0,1$  %-os, térben szigetelt esetben pedig  $P = 2,5$  %-os valószínűségi szinten volt szignifikáns a kapcsolat. Érzékelhető az is, hogy a szigetelés milyen mértékben befolyásolja a betegségek komplex kapcsolatát. Látható továbbá, hogy az 1. esetben amikor a csövet mesterségesen fertőzték, akkor a kapcsolat várható módon a csőpenész oldaláról lett dominánsabb. Normál esetekben viszont úgy tűnik, hogy NPK talajkezelés esetén a szárkorhadás a meghatározóbb. Következésképpen az ellenanyagproblémánál ez a betegség játszik elsődleges szerepet és ez intenzívebb kezelést igényel.

B.) Kórházban nyilvántartott, különböző életkorú és nemű cukorbetegek vérnyomását mérték. Ismert, hogy  $n$  számú cukorbeteg közül hány 0 - 20, 20 - 40, ... , 60 - 80 éves férfi és nő van. Továbbá, hogy közülük hányan volt magasvérnyomása.\* Kérdés, hogy van-e összefüggés és milyen korú és nemű betegeknél a cukorbetegség és a magasvérnyomás között?

A megoldás alapgondolata itt is az, hogy a tapasztalt gyakorisági eloszlások egy  $k \times v$  mezős kontingencia-táblába rendezhetők, melyeknél több osztály esetén végzett homogenitásvizsgálat  $\chi^2$ -próbával dönt az összefüggés szignifikanciájáról. Kimutatható, hogy a két betegség esetenként egymással szignifikáns kapcsolatban van.

Az előadás első részében ismertetett  $K$ -mérték nemcsak két egymással kölcsönhatásban lévő betegség jellemzésére használható, hanem tetszőleges két kvalitatív változó között fennálló sztochasztikus kapcsolat erősségének mérésére is.

---

\* A problémát Vincze János (Kolozsvár) vetette fel.

	I.	II.	III.	IV.	V.
2	184,12	102,24	84,04	66,02	36,47
P %	0,1	0,1	2,5	30	99
Szign.	Szign.	Szign.	Szign.	Szign.	nem szign.
cső K <sub>1</sub>	0,20	0,12	0,08	0,07	0,03
max szár K <sub>2</sub>	0,14	0,15	0,09	0,04	0,10
Kont.	0,11	0,08	0,07	0,06	0,05

1. táblázat: A szárcorhadás és csőpenész összefüggésének eredményei kukoricánál

Irodalom

- (1) Deuchler, G.: Über die Methoden der Korrelationsrechnung in der Padagogik und Psychologie. Z. Pad. Psychol. Exp. Pad., 15 (1914) s. 114, 145, 229.
- (2) Deuchler, G.: Über die Bestimmung einseitiger Abhängigkeit padagogisch - psychologischen Tatbestanden mit alternativer Variabilität. Z. Pad. Psychol. Exp. Pad., 16 (1915), 550-560.
- (3) Ezekiel, M. - Fox, K.: Korreláció és regresszióanalízis, Közg. Jogi (1970).
- (4) Hellmich, K.: Das paarige Vierfeldertafel-Mass K, Acta Albertina Ratisbonensia, 29 (1969) 87-99.
- (5) Hellmich, K.: Paarige Vierfeldertafel-Masse, Psychologie und Praxis, 1 (1971) 38-42.
- (6) Hellmich, K.: Körperbau und Gattenwahl, Biometische Z., 15, (1973) 255-258.
- (7) Kendall, M. G.: Rank Correlation Methods, Verlag Charles Griffin and Co., London, (1948)
- (8) Sváb, J.: Biometriai módszerek a kutatásban, Mezőg. Kiadó, Bp. (1973).
- (9) Weber, E.: Grundriss der biologischen Statistik, G. Fischer, Jena, (1961).