A MCNEMAR-PRÓBA NÉHÁNY ÁLTALÁNOSÍTÁSA

Írta: HAJTMAN BÉLA

Bevezetés

A *McNemar*-próba néven közismert nemparaméteres statisztikai eljárás önkontrollos (vagy ennek megfelelő) kísérleti elrendezések esetén alkalmazható olyankor, ha a vizsgált változónak csupán két értéke lehetséges (illetve, ha az eredeti változót dichotomizáljuk). A továbbiakban a változót a kísérleti személyektől kapott *válasz*nak fogjuk nevezni, értékeire pedig a pozitív (+) és negatív (−) megjelöléseket fogjuk használni.

A próba a különböző értékek gyakoriságainak segítségével azt igyekszik megállapítani, hogy valamely beavatkozás megváltoztatja-e a kapott válaszok megoszlását. Nevezetesen, azt a nullhipotézist vizsgálja, hogy az egyes személyektől kapott válasz a beavatkozás után is ugyanaz, mint előtte volt — és az ettől való eltérések csak véletlen ingadozásoknak tekinthetők. A válaszok gyakoriságát egy 2×2-es táblázatban lehet célszerűen elrendezni (1. táblázat), ahol $a+b+c+d=n$ a kísérleti személyek száma.

1. TÁBLÁZAT

		Utána	
		+	−
Előtte	+	a	b
	−	c	d

A *McNemar*-próba [1] a fenti nullhipotézist a

$$\chi^2 = \frac{(b-c)^2}{b+c} \tag{1}$$

kifejezés segítségével vizsgálja; ez közelítőleg 1 szabadságfokú χ^2-eloszlást követ, ha a nullhipotézisnek megfelelő $\frac{b+c}{2}$ várt gyakoriság értéke 5 vagy annál nagyobb. (Egyébként a próba nem alkalmazható.)

Figyelembe véve a kísérletezés során fellépő gyakorlati igényeket, számos általánosítás, illetve korrekció látszik szükségesnek ezzel a próbával kapcsolatban. Ilyen általánosításnak fogható fel COCHRAN azon próbája [2] is, mely — az önkontrollos elrendezésre jellemző két minta helyett — több összetartozó minta vizsgálatára alkalmas, ugyancsak dichotomizált változók esetében. Bár a kísérleti elrendezés az előbbinek nyilvánvalóan általánosítása, COCHRAN a problémát más oldalról közelítette meg, és megoldásának formája sem mutat első látásra rokonságot MCNEMAR próbájával. (Valójában két minta esetén ekvivalens a két próba.)

A következőkben néhány olyan általánosításról lesz szó (részben csak a probléma felvetésének formájában), melyek közvetlenül a *McNemar*-próbából indulnak ki és azt igyekeznek néhány szempontból kiterjeszteni.

Az arányos változás problémája

Kiindulva a *McNemar*-próbának megfelelő kísérleti elrendezésből (1. táblázat), először is annak nullhipotézisét tesszük kritika tárgyává. MCNEMAR azt a nullhipotézist állította fel, hogy a véletlen okozta változások egyforma arányban oszlanak meg a pozitívból negatívba, illetve negatívból pozitívba forduló válaszok között. Ennek megfelelően, adottnak véve az „összes megváltozások" számát, a χ^2-próba alapjául szolgáló várt gyakoriság $\frac{b+c}{2}$ volt.

Ez a nézet, véleményünk szerint, tarthatatlanná válik, ha a válaszok kiindulási megoszlása erősen aránytalan. Ha pl. jóval több személytől kapunk negatív választ, mint pozitívot (azaz ha $a+b \ll c+d$), a nullhipotézisnek megfelelő esetben — tehát hatástalanság esetén — várhatóan több negatív válasz változik „véletlenül" pozitívvá, mint fordítva. Ezt a tényt figyelembe véve úgy fogalmazhatjuk a nullhipotézist, hogy a véletlen okozta megváltozások az eredeti csoportok *ugyanakkora részét* érintik. A két várt gyakoriság — a b mellett álló β és a c mellett álló γ — arányára tehát

(2) $$\beta : \gamma = a+b : c+d$$

kell hogy teljesüljön. Figyelembe véve még a

$$\beta+\gamma = b+c$$

kikötést,

$$\beta = \frac{(a+b)\,(b+c)}{n}$$

és

$$\gamma = \frac{(c+d)\,(b+c)}{n}$$

adódik.

Behelyettesítve ezeket a két cellából számított (1 szabadságfokú) χ^2 képletébe, rövid számolás után a

(3) $$\chi^2 = \frac{(ac-bd)^2}{(a+b)(c+d)(b+c)}$$

összefüggést kapjuk. Ennek értéke számszerűleg megegyezik (1) értékével, ha $a+b=$ $=c+d$, egyébként azonban különbözik tőle. Könnyű olyan szituációkat találni, melyekben az egyik próba $\chi^2=0$-t ad, míg a másik erősen szignifikáns eredményt, más esetekben pedig a kimutatott változás ellentétes irányúnak adódik az (1), illetve (3) próba alapján számolva.

Az elmondottakból következik, hogy nem lenne sok értelme erőfüggvény segítségével keresni az igazolást, hogy melyik próba jobb: egyszerűen *mást* vizsgál az egyik, és mást a másik. A (3) próba bevezetésénél alkalmazott gyakorlati érvelést csak egyetlen szemponttal egészíteném ki. Az önkontrollos elrendezések esetén alkalmazott próbák (egymintás t-próba, *Wilcoxon*-próba, előjelpróba stb.) valamennyien invariánsak az időbeli megfordításra vonatkozólag. Legjobb tudomásom szerint egyetlen más próba sem használja ki azt az információt, amit az *előtte-utána* megkülönböztetés tartalmaz, vagyis ennek a viszonynak a tényleges *megfordíthatatlanságát*. Az pedig nyilvánvaló, hogy valamely probléma megoldására az a próba alkalmasabb, amelyik *többet* használ fel a feladat nyújtotta információk közül.

A (3) próba esetében is alkalmazható — természetesen kissé módosított formában — a *Yates*-féle folytonossági korrekció. Az (1) táblázat celláiban úgy kell az 1/2 értékeket hozzáadnunk, ill. levonnunk, hogy a várt gyakoriságok előállításához felhasznált (2) arány változatlan maradjon. Eszerint két-két egymás alatti cellában kell ugyanazt a korrekciót végezni.

Tegyük fel, hogy a válasz negatívba fordulását várjuk. Ehhez nem elég (de nem is kell), mint a *McNemar*-próba esetében, $b > c$ teljesülése: a kérdést a $b-\beta$ különbség előjele dönti el. Ennek értelmében a változás irányára vonatkozó feltevésünk ekvivalens azzal, hogy $bd-ac>0$ legyen. Ebből már következik, hogy az a és c értékeket kell növelnünk, a b-t és d-t pedig csökkentenünk 1/2-del.

Behelyettesítve az így módosított értékeket a (3) képletbe — és mindjárt az ellenkező feltevést, a pozitív irányú változást is figyelembe véve —, a (3) próba folytonossági korrekcióval módosított alakjára ezt kapjuk:

$$\chi^2 = \frac{(|ac-bd|-(n/2))^2}{(a+b)(c+d)(b+c)}. \tag{3)*}$$

A szakmai hatásosság figyelembevétele

Ezt a kérdést már Cochran [2] is érinti, sőt rámutat a megoldás nehézségeire is. Arról van szó, hogy a *McNemar*-próba (és ugyanígy Cochran általánosítása) teljesen figyelmen kívül hagyja az „előtte" és „utána" szituációban változatlan választ adó személyeket. Az (1) alatti képletből látszik, hogy a és d akármekkorák lehetnek, ez a próba eredményén nem változtat. Márpedig *szakmai* szempontból igen jelentős ezeknek az értékeknek a nagysága is. Tegyük fel, hogy bizonyos embereknek a véleményét akarjuk megváltoztatni, és azt szeretnénk, hogy akik negatív választ adtak, azok is igeneljék a szóban forgó kérdést. Ha a válaszok megváltozásából ezt a választott szinten szignifikánsan megerősíthetjük, látszólag igazoltuk meggyőzési módszerünk hatásosságát. Ehhez nem kell más, mint hogy $c > b$ legyen és (1) meghaladja a választott szignifikanciaszintnek megfelelő értéket. A feltételeknek megfelelő minimális számot választva ($b+c \geqq 10$), $c=9$, $b=1$ esetén (1) az 5%-os szinten szignifikáns lesz, még folytonossági korrekció bevezetése esetén is.[1] Világos azonban, hogy ennek

[1] A (3) próba esetében is alkalmazott *Yates*-féle korrekció (1) képletét így módosítja:

$$\chi^2 = \frac{(|b-c|-1)^2}{b+c}. \tag{1)*}$$

ellenére nem tekinthetjük — szakmai szempontból — hatásosnak a beavatkozásunkat, ha pl. 200 személy közül ($c+d=200$) került ki ez a 9 „meggyőzött". Ha viszont mindössze tizenketten voltak ($d=3$), akiket meg kellett győzni, eljárásunkkal meg lehetünk elégedve.

COCHRAN erre vonatkozóan azt mondja, hogy nem látja az utat, ahogyan ezt a problémát meg lehetne oldani anélkül, hogy a próba objektív voltát ne veszélyeztetnénk. Valóban így van, annyival inkább, mivel ez a „szakmai hatásosságra" vonatkozó követelmény problémánként más és más lehet. (Valamely gyógymód alkalmazandó akkor is, ha az — egyébként reménytelen — betegeknek csak egy kis százaléka is megenyhül, egy költséges reklámhadjárat viszont csak akkor kifizetődő, ha sok embert sikerül megnyerni a reklámozott ügynek.)

A probléma egyébként formailag könnyen megoldható. Ha az alternatív hipotézisben a változás irányát előre meghatározzuk (egyoldalú próbavégzés), a $\frac{c}{d}$ $\left(\text{vagy } \frac{b}{a}\right)$ arányra kell valamilyen kikötést tennünk, ha viszont ilyen előzetes hipotézisünk nincs (kétoldalú próba), a $\frac{b+c}{a+d}$ aránynak kell valamilyen előre meghatározott szintet meghaladnia. Elvileg megtehetjük azt is, hogy az előre kikötött (pl. legalább a populáció 40%-át érintő) változás teljesülését a binomiális eloszlás táblázata segítségével vizsgáljuk meg. Nem lenne azonban sok értelme ilyen pontos elemzést végezni egy végső soron *találomra* meghatározott, tetszőlegesen kitűzött „minimális hatásfok" ellenőrzése érdekében.

Végső soron abban lehet maradni, hogy esetenként szubjektív elbírálás tárgyává kell tenni, hogy a tapasztalt változások mennyisége elegendő-e ahhoz, hogy az alkalmazott beavatkozás hatásosságáról beszélhessünk. Ha a válaszunk igenlő, akkor láthatunk hozzá (a *McNemar*-próba vagy az előző szakaszban javasolt eljárás segítségével) annak eldöntéséhez, hogy a talált változás valóban alátámasztható-e az adatok alapján.

A kitűzött célt tehát nem értük el: a próbát nem sikerült megszabadítani a szakasz elején említett hiányosságától. Összegezésképp annyit mondhatunk, hogy szükségesnek tűnik a *McNemar*-próbát *kiegészíteni* egy előzetes, szubjektív „próbával", mivel önmagában — a feladatok gyakorlati célját tekintve — nem állja meg a helyét.[2]

A próba kiterjesztése kettőnél több értéket felvevő változókra

A kísérleti személyektől kapott válasz sok esetben finomabb osztályozást is megenged, mint az egyszerű dichotómiát. Ilyenkor is feltehetjük a kérdést, hogy a beavatkozás megváltoztatta-e a kapott válaszok arányait.

A feladatokat két részre oszthatjuk aszerint, hogy a lehetséges válaszok természetes sorrendbe (rangsorba) állíthatók-e vagy pedig nem. Nézzük először az utóbbi esetet. Ez az általánosabb, valójában azonban kevesebb érdekességet ígérő lehetőség. Egyszerűen általánosítva ugyanis a *McNemar*-próba gondolatmenetét,

[2] Hasonló probléma egyébként az előjelpróbával kapcsolatban is felléphet, de ott azért nem annyira szembetűnő, mert ritka az az eset, hogy a változatlan — tehát végső soron a mintából kihagyott — elemek száma számottevően nagy legyen.

2. TÁBLÁZAT

A satírozott mezőkben álló gyakoriságok a próba szempontjából közömbösek, ezért fel sem tüntettük őket.

		Utána		
		A	B	C
Előtte	A		n_1	n_2
	B	n_3		n_4
	C	n_5	n_6	

a feltételek olyan rendszeréhez jutunk, amelynek alapján valamennyi várt gyakoriság egymással egyenlő. A legegyszerűbb, három lehetséges választ tartalmazó esetben (l. a 2. táblázatot):

$$v = \frac{\sum_{i=1}^{6} n_i}{6}.$$

Az így kapott, 6 tagból álló χ^2 szabadságfoka 5. Szignifikáns eredményt kaphatunk, ha — mondjuk — n_2 és n_5 igen nagyok, a többi négy gyakoriság pedig kicsi. A gyakorlati feladat szempontjából azonban alig magyarázható az olyan beavatkozás, mely az A válaszokat C válasszá változtatja, ugyanakkor a C válaszokból A válaszokat csinál.[3]

Ami a másik lehetőséget illeti, a rangsorba állítható válaszok esetét, ez gyakorlati szempontból sokkal fontosabbnak tűnik. A *McNemar*-próba feladatának megfogalmazásában *pozitív* és *negatív válaszokat* emlegettünk; ezzel mintegy elköteleztük a próbát az *attitűdök* vizsgálatára. Márpedig az attitűdök rendezhető (sorbaállítható) értékű változót képeznek, és ritkán szorítkoznak csupán két értékre. A 3. táblázat az 5 lehetséges értéket felvevő válasz esetében mutatja a gyakoriságok elrendezésének sémáját. Ebben az esetben a válasz a leghatározottabb igenléstől (+ +) a közömbösségen át (0) az erős ellenzésig (— —) terjed. Máskor a skála pl. így fogalmazható: nagyon szereti, szereti, közömbös, nem szereti, ki nem állhatja.

Az attitűdök azonban csak egyik (talán a legtipikusabb és leggyakoribb, de nem egyetlen) példája a rangsorolható válaszoknak. Az ismertetendő eljárás nem szorítkozik az attitűdök vizsgálatára — épp úgy, ahogy nem korlátozódik az öt különböző értéket felvevő változókra sem.

[3] Pszichológiai problémáknál elképzelhető hasonló jelenség; erre vonatkozó — bár fiktív — példát említek a [3] könyv 329. oldalán. Mindenesetre pozitívumként kell elkönyvelni azt, hogy a fent vázolt próba *képes* az ilyen jellegű változások kimutatására.

3. TÁBLÁZAT

		Utána				
		$++$	$+$	0	$-$	$--$
Előtte	$++$	n_{11}	n_{12}	$\cdots$	$\cdots$	n_{15}
	$+$	n_{21}	n_{22}	$\cdots$	$\cdots$	n_{25}
	0	$\vdots$				
	$-$	$\vdots$				
	$--$	n_{51}	n_{52}	$\cdots$	$\cdots$	n_{55}

Általánosságban, tekintsünk egy önkontrollos (vagy azzal egyenértékű, páronként összetartozó elemekből álló) elrendezést, melyben a változó (a válasz) k lehetséges értéket vesz fel ($k>2$). A két mintában — a bevezetett szóhasználattal a beavatkozás *előtt* és *után* — kapott válaszokat egy $k\times k$ mezős táblázatban helyezhetjük el a 3. táblázat mintájára, ahol

$$\sum_i \sum_j n_{ij} = N$$

a minta elemszáma. Szükségünk lesz még a sorösszegekre is:

$$\sum_j n_{ij} = N_i$$

az i-edik sor elemeinek összege, az „előtte" i-edik választ adó személyek száma.

Egy ilyen elrendezésben a változás „mértékét" a cella két indexének eltérése jellemzi: az $i-j$ szám megmutatja, hogy a cellához tartozó n_{ij} számú személy válasza milyen mértékben változott meg a beavatkozás hatására. Ha $i-j$ pozitív, a válasz „pozitívabb" lett (pl. kevésbé elutasító, közömbös vagy igenlő, ha pozitív volt, akkor még pozitívabb), ha negatív, „negatívabb". (L. a 3. táblázatot.) A főátlóban azok állnak, akik nem változtatták meg véleményüket; valóban, ezekben a cellákban $i-j=0$ adódik.

Vizsgáljuk azt a nullhipotézist, hogy a beavatkozás nem változtatja meg a kapott válaszok arányait, illetve, hogy az ilyen változások csupán véletlen ingadozásoknak tekinthetők. A hipotézis ellenőrzésére a főátlón kívüli $k(k-1)$ cellára vonatkozó tagok összegeként előállított χ^2 mennyiség szolgál; feladatunk az ennek felírásához szükséges várt gyakoriságok előállítása. Jelöljük ezeket (a már előbb is alkalmazott görög betűs jelölés analógiájára) ν_{ij}-vel.

Elsőnek a (3) alatti próba általánosítását állítjuk elő. Ennek megfelelően, az egyes sorokban az összes változások (megváltozott vélemények) számát a nullhipotézis fennállása esetén a sor létszámával arányosnak tételezzük fel:

$$\frac{\sum\limits_{j\neq i_1} \nu_{i_1 j}}{\sum\limits_{j\neq i_2} \nu_{i_2 j}} = \frac{N_{i_1}}{N_{i_2}}, \quad i_1, i_2 = 1, 2, \ldots, k. \tag{4}$$

Az *összes* megváltozások számát, mint mindig, adottnak vesszük:

$$\sum_{i \neq j}\sum v_{ij} = \sum_{i \neq j}\sum n_{ij}. \tag{5}$$

Ennyi azonban nem elég a várt gyakoriságok előállításához. (4) és (5) alapján megkapható a várt gyakoriságok összege valamennyi sorban, de kellene valamilyen elv, ami szerint a cellákban „szétosztjuk" őket.

Térjünk vissza a nullhipotézishez. Kézenfekvő az a megállapítás, hogy ha a változások csupán a véletlennek tulajdoníthatók, a nagy változások jóval ritkábbak, mint a kicsik. Mivel pedig a megváltozásban mintegy „hibát" látunk a feltételezett változatlansággal szemben,[4] természetes következtetés, hogy a megváltozások előfordulása a normális eloszlás (a „hibafüggvény") törvényeit követi. A gyakorlatban ez azt jelenti, hogy pl. az első sor várt gyakoriságai, $v_{12}, v_{13}, \ldots, v_{1k}$ úgy aránylanak egymáshoz, mint a normális eloszlás megfelelő helyein vett valószínűségek. A problémát éppen ezeknek a „megfelelő helyeknek" a kijelölése okozza. Meglehetős önkényesen úgy jártunk el, hogy a normális eloszlás görbéjét a kétoldalú 5%-os szignifikanciaszintnek megfelelő értéknél ($\pm 1{,}960$-nál) csonkítottuk, majd a maradékot $2k-1$ *egyenlő részre* osztottuk. Így a középső, „legvalószínűbb" szakasztól jobbra is, balra is megkaptuk a változásoknak megfelelő $k-1$ szakaszt. A normális eloszlás bármely eléggé részletes táblázatából könnyű meghatározni az ezekhez a szakaszokhoz tartozó valószínűségeket. Minket azonban nem maguk a valószínűségek érdekelnek, hanem azok egymáshoz viszonyított arányai: ezek az arányok jellemzik — modellünk értelmében — az egy sorba kerülő v_{ij} várt gyakoriságok megoszlását.

A várt gyakoriságokat meghatározó egyenletek választhatók úgy, hogy ne legyen szükség valamennyi arány meghatározására. Ha minden várt gyakoriságot a főátló melletihez viszonyítunk, $k-2$ arány megadása elegendő. A konkrét munka megkönnyítése érdekében $k \leq 10$ esetére táblázatba foglaltuk ezeket a P_j^k-val jelzett arányszámokat (4. táblázat); a k felső index a változó különböző értékeinek számát mutatja (ez a *feladat* paramétere), j pedig a főátlótól való távolságot ($j = 1, 2, \ldots, k-1$). A normális eloszlás szimmetriája miatt csak pozitív j-értékeket vettünk figyelembe.

A táblázatbeli értékek segítségével az egyes sorokban álló (és összegükben már meghatározott) v_{ij} várt gyakoriságok így jellemezhetők:

$$(6) \qquad \begin{aligned} v_{ij} &= P_{j-i}^k v_{i,i+1}, \quad \text{ha } j > i; \\ v_{ij} &= P_{i-j}^k v_{i,i-1}, \quad \text{ha } j < i. \end{aligned}$$

Az elmondottakból következik, hogy minden esetben igaz, hogy $v_{i,i+1} = v_{i,i-1}$ ($i = 2, \ldots, k-1$).

A felállított modell önkényessége nem az 5%-os (meglehetősen konvencionális) csonkításban, hanem az ekvidisztans felosztásban van. Amíg mérési eredmények ilyen úton történő vizsgálatáról van szó, tökéletesen helytálló az eljárásunk, ha

[4] Valóban hibáról van szó, ha ezen itt nem is (vagy nem kizárólag) a szokásos értelemben vett mérési hibát értjük. Ha igaz a nullhipotézis, a megkérdezett személynek semmi oka, hogy válaszát megváltoztassa. Ha mégis ezt teszi, ez valamilyen belső határozatlanság — a saját véleménye pontos megállapításának hibája! — miatt van.

4. TÁBLÁZAT

A P_j^k értékek táblázata

k \ j	1	2	3	4	5	6	7	8	9
2	1	—	—	—	—	—	—	—	—
3	1	0,416	—	—	—	—	—	—	—
4	1	0,633	0,295	—	—	—	—	—	—
5	1	0,756	0,474	0,246	—	—	—	—	—
6	1	0,828	0,605	0,390	0,221	—	—	—	—
7	1	0,873	0,697	0,508	0,339	0,206	—	—	—
8	1	0,903	0,762	0,600	0,443	0,305	0,196	—	—
9	1	0,924	0,809	0,672	0,530	0,396	0,281	0,189	—
10	1	0,938	0,844	0,727	0,601	0,476	0,361	0,263	0,183

egyenletes osztályszélességet választottunk a gyakoriságokon keresztül történő vizsgálatra való áttéréskor. Megállapítható vagy rangsorolható változók (pl. attitűdök) esetében azonban az egyenletes beosztás erősen támadható. Az attitűdök (és hasonló változók) mérési skálára való áttételekor ugyan rendszerint a normális eloszlást választják a skála meghatározásához (l. pl. a [4] és [5] monográfiákat), a skálázások konkrét elvégzésekor azonban rendszerint kiderül, hogy a skála — a normális eloszlásra „illesztve" — nem ekvidisztans. Mivel azonban egy-egy konkrét próbavégzés esetén a skálázási eljárás végrehajtására aligha van mód (rendszerint a minta nagysága sem elegendő ehhez), csak azokban az esetekben nem ajánlható a fent leírt eljárás alkalmazása, ahol a változó beosztásának „nem egyenletes" volta plauzibilis.[5] A 3. táblázaton jelzett attitűd-skála esetében például aligha van durva eltérés az ekvidisztanciától.

Valószínűleg jobb, de az ajánlottnál lényegesen bonyolultabb volna a normális eloszláshoz való illesztés megvalósítására egy olyan eljárás, mely a főátlóban álló gyakoriságokat is figyelembe veszi. Felvéve ismét egy $2k-1$ szakaszos ekvidisztans beosztást (mely ebben az esetben tetszőleges sűrűségű lehet), a változatlan választ adók gyakorisága segítségével becsüljük meg annak a normális eloszlásnak a *szórását,* amelyet erre a beosztásra illesztünk. Ha a főátló elemei eléggé egyformák, egységesen határozhatjuk meg valamennyi sorra a várt gyakoriságok egymáshoz viszonyított arányát. Ha azonban soronként végezzük, célszerű a (4) feltétel helyett soronként kikötni a

$$\sum_{j \neq i} \nu_{ij} = \sum_{j \neq i} n_{ij} \quad i = 1, 2, \ldots, k$$

[5] Elvileg járható a következő út. A mintánk eloszlásának a normális eloszlás skáláján megfelelő pontokat meghatározzuk az $N_1, N_2, \ldots, N_k$ számok alapján. Az így kapott szakaszokból azután minden sorban külön-külön meghatározzuk a normális eloszlásnak azt a beosztását, amiből a $P_j^{k,i}$ arányok adódnak. (Itt tehát nem *feladatonként,* hanem *soronként* más arányokkal kell dolgoznunk.) Közben meg kell oldanunk azt a problémát, hogy mekkora szakaszt „tartunk fenn" a változatlanok számára. Ezt ismét végezhetjük soronkénti becsléssel, de felhasználhatjuk a $\frac{\sum n_{ii}}{N}$ arányt is. Minderre azonban csak akkor kerülhet sor, ha az N_i számok által meghatározott eloszlás egyáltalán illeszthető a normálishoz. Nem ritka azonban az az eset, hogy ez egy *U*-alakú (bimodális) eloszlás: a kérdéssel kapcsolatban a populáció két szélsőséges táborra oszlik.

egyenlőség teljesülését. Ezek az újabb megszorítások azonban erősen csökkentik a próba szabadságfokát. (A becslések nem vesznek el a szabadságfokból, tekintve, hogy azokra az egyébként föl nem használt n_{ii} elemeket vesszük csak igénybe.)

A (4)—(6) feltételek felhasználásával a főátló melletti várt gyakoriságok így állíthatók elő:

$$\nu_{i,i\pm 1} = \frac{N_i \sum\sum_{i \neq j} n_{ij}}{N\left(\sum_{j=1}^{k-i} P_j^k + \sum_{j=1}^{i-1} P_j^k\right)}. \tag{7}$$

A többi várt gyakoriság már könnyen kapható (6) segítségével.

Mivel a várt gyakoriságok előállításához felhasznált feltételek közül csak (5) volt az, amelyik számszerű megkötést tartalmazott rájuk vonatkozólag (a többi csak egymáshoz viszonyított *arányukat* rögzítette), a kapott χ^2 szabadságfoka k^2-k-1 lesz.

Egészen hasonló módon kaphattuk volna az (1) próba általánosítását, a (4) kikötés helyett az egyenletesen elosztott változást feltételező

$$\sum_{j \neq i} \nu_{ij} = \frac{\sum\sum_{i \neq j} n_{ij}}{k} \qquad (i = 1, 2, \ldots, k)$$

feltételt használva. Meg kell azonban mondani, hogy ez a próba itt sokkal kevésbé természetes, mint a $k=2$ esetben. Ha viszont kiegészítjük egy, az *oszlopok* egyenletes eloszlására vonatkozó hasonló feltétellel, ez már $k=3$ esetben sem egyeztethető össze a normális eloszlást feltételező modellel. Mivel pedig aszimmetrikus eloszlást észszerűtlen volna feltételezni, egyetlen lehetőség az egyforma nagy várt gyakoriságok megállapítása:

$$\nu_{ij} = \frac{\sum\sum_{i \neq j} n_{ij}}{k(k-1)} \quad i \neq j. \tag{8}$$

Ezzel visszajutottunk ahhoz a próbához, amit a szakasz elején, a nem rendezhető változók esetében kaptunk. Valóban: itt sehol nem használjuk ki a rendezettséget — és az ezzel járó előnyöket.

A fönt vázolt valamennyi eljárás ekvivalens $k=2$ esetén az (1), ill. (3) próbákkal. Van azonban arra is lehetőség, hogy *közvetlenül* alkalmazzuk a *McNemar*-próbát $k>2$ esetén, ha a változó értékei rendezhetők. Visszavezetve ugyanis a 3. táblázathoz hasonló (de akármekkora) elrendezéseket az 1. táblázatra, értelemszerűen igaz

$$\sum\sum_{i<j} n_{ij} = b$$

és

$$\sum\sum_{i>j} n_{ij} = c.$$

A (3) próbát nem alkalmazhatjuk így, mert ahhoz az a és d mennyiségekre is szükség van, azokat azonban nem tudjuk a $k>2$ esetből kiindulva definiálni.

Talán nem szükséges hangsúlyozni, hogy a legutóbb leírt próba $k>2$ esetén *nem ekvivalens* a (8) alatti várt gyakoriságokkal megadott próbával. Az utóbbi a változás „tendenciáján" (irányán) kívül a cellákban történő egyenletes eloszlásra is érzékeny; más kérdés, hogy ezt az egyenletességet, illetve az ettől való eltérést sokszor nehéz interpretálni.

A 3. táblázat típusába tartozó gyakorisági elrendezések vizsgálatára egyéb módok is lehetségesek. Meg kell itt említenünk BOWKER [6] próbáját, aki a kontingenciatáblázat szimmetriájának vizsgálatára dolgozta ki a

$$\chi^2 = \sum_{i>j}\sum \frac{(n_{ij}-n_{ji})^2}{n_{ij}+n_{ji}} \tag{9}$$

próbát. $\left(\text{Szabadságfok: } \frac{k(k-1)}{2}.\right)$ Ezzel ismét az a probléma, hogy nehéz olyan gyakorlati feladatot találni, amelynek megfelel ez a kérdésfeltevés. A válaszok megváltozásának önkontrollos elrendezés segítségével történő vizsgálatára a (9) próba alkalmatlan.

Befejezésül csak megemlítenénk egy, az eddigiektől eltérő lehetőséget a várt gyakoriságok előállítására a $k>2$ esetben. Hasonlóan a *McNemar*-próbához, azt kötjük ki, hogy

$$\sum_{i<j}\sum v_{ij} = \sum_{i>j}\sum v_{ij} = \frac{\sum\limits_{i\neq j}\sum n_{ij}}{2} \tag{10}$$

legyen, továbbá, hogy a várt gyakoriságokat úgy kell megállapitani, hogy a χ^2 értéke minimális legyen. (Így valóban csak a (10) feltételt használjuk fel: csak akkor fogunk tudni kimondani szignifikanciát, ha *minden* olyan χ^2-próba, melyet (10)-nek megfelelő módon konstruáltunk, szignifikáns.)

A számítások bonyolultsága miatt ehelyett azt a — vele nyilván nem egyenértékű — problémát oldottuk meg, hogy mik lesznek a várt gyakoriságok, ha (10) feltevése mellett még azt követeljük meg, hogy a

$$\sum_{i\neq j}\sum (n_{ij}-v_{ij})^2$$

összeg legyen minimális. Ekkor $i<j$ esetére a következő összefüggést kaptuk:

$$v_{ij} = n_{ij} + \frac{\sum\limits_{i>j}\sum n_{ij} - \sum\limits_{i<j}\sum n_{ij}}{k(k-1)}. \tag{11}$$

($i>j$ esetén értelemszerűen úgy módosul a képlet, hogy a tört előjele ellenkezőjére változik.) A szabadságfok ilyenkor k^2-k-2 lesz.

A (11) által megadott várt gyakoriság zérus, sőt negatív is lehet; különösen fontos tehát, hogy a χ^2-próba alkalmazhatóságának a várt gyakoriságokra vonatkozó általánosan elfogadott feltételeit figyelembe vegyük. (A várt gyakoriságoknak 5-nél nagyobbnak kell lenniök, illetve egyes szerzők szerint néhány cellában 1 és 5 közötti értékeket is meg lehet engedni.) A feltétel csak akkor tud teljesülni, ha a változás irányába eső gyakoriságok (pl. a válaszok „negatívabbá válása" esetén a jobb felső háromszög n_{ij}-i) viszonylag egyenletesen nagyok. Ez azonban alkalmasint olyan ritkán megvalósuló feltétel, hogy ez az utolsóként említett próba ezáltal elveszti minden gyakorlati jelentőségét.

IRODALOM

[1] McNemar, Q.: Note on sampling error of the difference between correlated proportions or percentages, *Psychometrika* **12** (1947), 153—157.
[2] Cochran, W. G.: The comparison of percentages in matched samples, *Biometrika* **37** (1950), 256—266.
[3] Hajtman, B.: *Bevezetés a matematikai statisztikába, pszichológusok számára.* Akadémiai Kiadó, Budapest (1968).
[4] Edwards, A. L.: *Techniques of Attitude Scale Construction.* Appleton—Century—Crofts, New York (1957).
[5] Torgerson, W. S.: *Theory and Methods of Scaling.* Wiley, New York (1958).
[6] Bowker, A. H.: A test for symmetry in contingency tables, *J. Amer. Stat. Ass.* **43** (1948), 572—574.

(Beérkezett: 1969. VII. 2.)

SOME GENERALIZATIONS OF MCNEMAR'S TEST

by

B. Hajtman

Summary

The well-known *McNemar* test (1) is criticized as regards its null-hypothesis. It is not obvious (see Table 1) that under null-hypothesis $b=c$ holds; the changes are rather proportional to the original contingencies (the marginal distribution "before"). Under this assumption we get the test (3).

After arising an unsolved problem (to incorporate the "unchanged" persons a and d in the test) an attempt is made to generalize both tests (1) and (3) for non-dichotomized answers (for $k>2$ instead of McNemar's case $k=2$). Among several possibilities the test characterized by (4)—(7) seems to be the most appropriate. This test is in particular intended to investigate the changes in attitudes. The P_j^k coefficients in Table 4 are calculated by the aid of areas under the normal curve.