

KÉT PSZICHOLÓGIAI POPULÁCIÓ SZTOCHASZTIKUS EGYENLŐSÉGÉNEK ELLENŐRZÉSÉRE ALKALMAS STATISZTIKAI PRÓBÁK ÖSSZEHASONLÍTÓ VIZSGÁLATA*

VARGHA ANDRÁS

ELTE, Bölcsészettudományi Kar, Kísérleti Pszichológia Tanszék
E-mail: vargha@izabell.elte.hu

A jelen tanulmányban a sztochasztikus egyenlőség ellenőrzésére alkalmas hat statisztikai próbát hasonlítottunk össze számítógépes szimulációval az érvényesség és a hatékonyság kritériuma szempontjából. Két populációt akkor mondunk sztochasztikusan egyenlőnek valamely X változó tekintetében, ha véletlenszerűen kiválasztva egy-egy X-értéket a két populációból, az elsőből kiválasztott érték ugyanakkora eséllyel lesz nagyobb a második kiválasztottnál, mint kisebb.

A szimulációban széles tartományban variáltuk az eloszlások ferdeségét és csúcsosságát, valamint a szórásheterogenitás mértékét. Vizsgáltunk kicsi és közepes nagyságú, illetve egyenlő és különböző elemszámú mintákat. A szimulációba a korábban már mások által is javasolt próbák (rang t, rang Welch, Fligner–Policello, Cliff) mellett elméleti megfontolások alapján két új próbát (FPW és FPCW) is bevontunk.

A szimulációs vizsgálatok arra az érdekes eredményre vezettek, hogy az újonnan javasolt két próba, FPW és FPCW az érvényesség tekintetében sokkal megbízhatóbb eljárásnak bizonyult, mint a többiek, miközben az erő tekintetében nem tapasztaltunk számottevő különbséget közöttük. Különösen FPW jeleskedett azzal, hogy I. fajta hibája sosem tért el számottevően a névleges szinttől. Közepes nagyságú minták esetén FPCW FPW-vel egyenértékű eljárás benyomását keltette.

Kulcsszavak: csoportösszehasonlítás, sztochasztikus egyenlőség, sztochasztikus különbség, valószínűségi fölény mutatója, rang Welch-próba, Fligner–Policello-próba, Fligner–Policello-próba Cliff-féle módosítása, FPW-próba, FPCW-próba

* A tanulmány megírásához nagy segítséget nyújtott a szerző 1997-ben elnyert Széchenyi Professzori Ösztöndíja, a T032157 számú OTKA és a 0194/2000 számú FKFP pályázat, valamint a Research Support Scheme of the Open Society Support Foundation 584/1998 számú pályázata.

BEVEZETÉS

Ha van egy kvantitatív jellegű változónk és szakmai kérdésünk az, hogy e változó értékei ugyanakkor-e két pszichológiai populációban, akkor független adatminták esetén a legerjedtebb megoldás az átlagok és szórások segítségével operáló kétmintás t-próba vagy annak robusztus változata, a Welch-féle d-próba (vö. HAJTMAN, 1968, 159–167; VARGHA, 2000, 231–239). Ez esetben akkor tekintünk két populációt ugyanakkorának, ha a vizsgált változó a két populációban ugyanolyan átlagú.

Az átlag azonban nem mindig tükrözi jól a változó értékeinek nagyságintjét. Ha a változó erősen ferde eloszlású (ilyen például a legtöbb reakcióidő változó), akkor az extrém értékek „maguk felé húzzák” az átlagot, azt eredményezve, hogy az átlag alatti és fölötti értékkel jellemezhető személyek aránya esetenként számottevően eltér egymástól, illetve a szimmetrikus eloszlásokra jellemző 50-50%-tól.

A medián olyan középérték, mely folytonos eloszlású változók esetén mindig 50-50%-os arányban osztja két részre (a mediánnál kisebb, illetve nagyobb értékűek alpopulációjára) a vizsgált összességet. A mediánnal kapcsolatban viszont problematikus, hogy ez a szép tulajdonsága csak folytonos eloszlásokra érvényes, miközben a pszichológiai empirikus vizsgálatok változói többnyire diszkrét (vö. MICCERI, 1989).

Újabbban kezd elterjedni a trimmelt átlag (trimmed mean) használata. Ez olyan robusztus középérték, mely az átlagnál kevésbé érzékeny a szélsőségesen nagy vagy kicsi értékek előfordulására. Képzésekor ugyanis levágjuk az eloszlás bizonyos nagyságú – legtöbbször 10 vagy 20%-ot kitevő – szélét (balról és jobbról egyaránt), s a maradék tartományba eső értékekből számítunk hagyományos módon átlagot (lásd részletesebben WILCOX, 1996, 15). Ez a mutató olyan középérték, melyet kevésbé befolyásolnak az egyedi, ritka, az átlagostól erősen elütő, kiugró értékek, s emiatt az adatok legbelső centrumában kell elhelyezkednie. Problémás azonban vele kapcsolatban definíciójának túlzott rugalmassága. Mekkora nagyságú részt vágunk le az eloszlás két szélén: 5, 10 vagy 20%-ot? Ha az eloszlás erősen ferde, akkor a két oldalról ugyanakkora vagy eltérő nagyságú részt vágunk-e le (pl. 5-5%-ot vagy inkább 3-7%-ot)?

Minél több mutató terjed el a változók nagyságintjének mérésére, annál inkább megnő az esélye a szakmai félreértéseknek. Egy olyan kijelentés például, hogy egy változó – mondjuk egy egyszerű reakcióidő – nagyságintje a kísérleti és a kontroll helyzetben ugyanakkora, vagyis hogy az adott helyzetben a kísérleti hatás nulla, egyaránt megfogalmazódhat hagyományos átlagokkal, trimmelt átlagokkal, illetve mediánokkal végzett nem szignifikáns statisztikai összehasonlítás nyomán. Tekintve azonban, hogy e különböző középértékek elméletileg különböző állapotokat is tükrözhetnek, a velük végzett statisztikai próbák esetenként teljesen eltérő eredményre vezethetnek (lásd pl. WILCOX, 1996, 153–154).

A populációk középértékek segítségével történő összehasonlításai mind azt a gondolatmenetet követik, hogy először kiválasztanak egy speciális középérték típusát (átlag, trimmelt átlag, medián stb.), majd a populációk ugyanakkorosságát e középértékek egyenlőségével definiálják, vagyis két populáció közül valamely X

változó tekintetében azt tekintik nagyobbak, amelyeknek az adott középértéke (pl. az átlaga) a populációban nagyobb. A populációk nagyság szerinti összehasonlítását azonban nemcsak közvetve, bizonyos középértékeik útján, hanem közvetlenül, értékeik direkt összehasonlításával is el lehet végezni. Egy előző tanulmányunkban (VARGHA, 1999a) ismertettük erre a célra a valószínűségi fölény A mutatóját, mely a statisztikában kevésbé járatosak számára is könnyen érthető, s amely két összehasonlítandó populáció esetén (jelölje őket mondjuk Pop1 és Pop2) lényegében egy százalékos skálán méri Pop1 dominanciáját Pop2-vel szemben. A értéke 0 (= 0%) és 1 (= 100%) között változhat. Az $A = 1$ érték Pop1 teljes fölényét jelzi Pop2-vel szemben (ilyenkor bármely Pop1-beli X-adat nagyobb bármely Pop2-beli X-adatnál), $A = 0$ pedig Pop2 teljes fölényét Pop1-gyel szemben (bármely Pop2-beli X-adat nagyobb bármely Pop1-beli X-adatnál). $A = 0,5$ esetén azt mondjuk, hogy a két populáció az X változó tekintetében sztochasztikusan egyenlő. Ez akkor fordul elő, ha a két populációból véletlenszerűen kiválasztva egy-egy értéket (jelölje ezeket mondjuk X_1 és X_2), az elsőből kiválasztott ugyanolyan valószínűséggel lesz nagyobb a másodikból kiválasztottnál, mint fordítva:

$$P(X_1 > X_2) = P(X_1 < X_2). \quad (1)$$

A valószínűségi fölény A mutatója bármely olyan X változó esetén értelmezhető, amelynek értékskálája legalább ordinális, tehát amelynek az értékei valamely értelmes szakmai szempont alapján nagyság szerint összehasonlíthatók. Nem szükséges sem a normalitás, sem a kvantitatív korlátozó feltétele: csak az ordinalitás. E mutatóval kapcsolatos további részleteket illetően lásd VARGHA (1999a).

Ezek után fontos kérdés, hogy hogyan lehet eldönteni két populációról, hogy sztochasztikusan egyenlők-e. Idézett tanulmányunkban bebizonyítottuk, hogy két populáció (1) formulával definiált sztochasztikus egyenlősége matematikailag egyenértékű egy másik összefüggéssel, amelyhez a következőképpen jutunk. Válasszunk ki véletlenszerűen és egymástól függetlenül egy-egy mintát a Pop1 és a Pop2 populációból, majd rangsoroljuk a kapott adatokat nagyság szerint úgy, ahogy az a jól ismert Mann-Whitney rangsorolós eljárásban is történik (lásd pl. HAJTMAN, 1968, 348; VARGHA, 2000, 259). Említett cikkünkben azt igazoltuk, hogy a két populáció sztochasztikus egyenlősége akkor és csakis akkor teljesül, ha a Pop1-ből véletlenszerűen kiválasztott minta rangszámainak az elméleti átlaga megegyezik a másodikból kiválasztott véletlen minta elméleti átlagával, ha tehát a rangszámok populációi ugyanakkora elméleti átlagúak.

Ennek az összefüggésnek az a jelentősége, hogy utat mutat a sztochasztikus egyenlőség hipotézisének statisztikai ellenőrzéséhez. Ugyanis két elméleti átlag egyenlőségének ellenőrzésére számos statisztikai próba áll rendelkezésre. A legismertebb közülük a jelen cikk elején is említett kétmintás t-próba. Ha tehát a kétmintás t-próbát a rangsorolás után kapott rangszámokon hajtjuk végre, akkor ezzel a sztochasztikus egyenlőséget tesztelhetjük. Megjegyezzük, hogy ez a rangszámokon végzett t-próba (nevezzük ezt rang t, rövidítve r_t -próbának) lényegében ugyanaz, mint a Mann-Whitney-próba nagy mintákra vonatkozó eljárása (lásd pl. CONOVER, IMAN, 1981; ZIMMERMAN, ZUMBO, 1993a, 1993b; MCKEAN, VIDMAR, 1994). Ismeretes azonban, hogy a kétmintás t-próbának szigorú alkalmazási felté-

tele a szóráshomogenitás. A rang t-próba és a Mann-Whitney-próba közös alkalmazási feltétele eszerint, hogy az elméleti szórások (a rangszámok esetében az elméleti rangszórások) egyezzenek meg. Tekintve azonban, hogy az elméleti szórások egyenlősége a gyakorlatban sokszor nem teljesül, illetve alkalmas próbák híján nem lehet róla megbizonyosodni, olyan – ún. *robustus* – statisztikai próbákra lenne szükség, amelyek ezt a szigorú korlátozást nem igénylik. Előző cikkünkben (VARGHA, 1999a) három ilyen robustus eljárást: a rang Welch-próbát (ZIMMERMAN, ZUMBO, 1992, 1993a), a Fligner–Policello-próbát (FLIGNER, POLICELLO, 1981) és ez utóbbinak a Cliff által módosított változatát (CLIFF, 1993) javasoltuk a sztochasztikus egyenlőség általános, feltételektől mentes ellenőrzésére két független minta esetén.

A jelen tanulmány fő célja e robustus próbák alkalmasságának (érvényességének és erejének) vizsgálata számítógépes szimulációval. A tanulmány 1. pontjában e próbákkal kapcsolatos témánkba vágó irodalmi adatokat tekintjük át, s javasolunk két újabb eljárásváltozatot a sztochasztikus egyenlőség ellenőrzésére.

A 2. pontban a szimuláció modelljét és technikai részleteit írjuk le.

A 3. pontban az egyenlő és a különböző mintaelemszámokra vonatkozóan elemezzük a szimuláció érvényességre vonatkozó eredményeit (I. fajta hibák), valamint azt, hogy az érvényesnek talált próbák ereje hogyan alakul, vagyis hogy különböző feltételek mellett milyen érzékenységgel képesek a sztochasztikus egyenlőség állapotát jelezni.

1. A VIZSGÁLT TESZTELJÁRÁSOK

A jelen tanulmányban ismertetésre kerülő számítógépes szimulációs vizsgálatoknak az volt a kitűzött célja, hogy empirikus információhoz jussunk a sztochasztikus egyenlőség különböző statisztikai próbáinak érvényességével kapcsolatban. A fő kérdés az, hogyan alakul ezeknek az eljárásoknak az elsőfajú hibája és ereje különböző eloszlások, elemszámok és elemszám-szórás kombinációk mellett. A robustus próbáktól azt várjuk el, hogy tényleges elsőfajú hibájuk a sztochasztikus egyenlőség nullhipotézisének vizsgálata esetén ne térjen el számottevően a névleges szinttől akkor se, ha az elméleti szórások jelentősen különböznek, és pedig a függő változó eloszlásának széles tartományában.

Különösen fontosak számunkra a nem szimmetrikus eloszlások generálásával végzett szimulációk, mert ez esetben a sztochasztikus egyenlőség általában más viszonylatot jelent a két populáció között, mint az elméleti átlagok vagy mediánok egyenlősége (lásd erről részletesebben VARGHA, DELANEY, 1998), de erre vonatkozóan nem találunk behatóbb érvényességi vizsgálatokat a statisztikai irodalomban. Ugyanakkor van néhány olyan tanulmány, amely a szimmetrikus eloszlások családján belül (ez esetben a sztochasztikus egyenlőség ekvivalens az elméleti átlagok egyenlőségével, folytonos esetben pedig még a mediánok egyenlőségével is) elemezte a rang Welch- (rövidítve rW) és a Fligner–Policello- (rövidítve FP) próba érvényességét az elméleti átlagok vagy a mediánok egyenlősége nullhipotézis rögzítése mellett.

ZIMMERMAN és ZUMBO (1993a) normális eloszlású függő változókkal kimutatta, hogy az rW-próba megtartja érvényességét az I. fajta hiba tekintetében jelentősen különböző szórások esetén is, egyenlő és különböző mintaelemszámok mellett egyaránt. E szerzők hasonló eredményeket kaptak enyhén megemelkedett elsőfajú hibával néhány más folytonos szimmetrikus eloszlás (Cauchy, Laplace, egyenletes, kevert normális, kevert egyenletes) esetén is (ZIMMERMAN, ZUMBO, 1992, 1993a). Zimmerman és Zumbo ugyanezen szimulációs vizsgálatokat néhány nem szimmetrikus eloszlással (exponenciális, lognormális stb.) is elvégezte, de erre vonatkozó eredményeik sajnálatosan nem értelmezhetők a sztochasztikus egyenlőség szempontjából. Ennek az az oka, hogy amikor a várható értékek egyenlőségének megtartása mellett ezeket a kezdetben azonos ferde eloszlásokat lineáris transzformációval különböző szórásúvá transzformálták a két mintában, a sztochasztikus egyenlőséget „elrontották”, s emiatt a szimulációban kapott szignifikáns eredmények aránya nem az elsőfajú hibát, hanem a próba erejét becsülte (vö. pl. ZUMBO, COULOMBE, 1997, 143. oldal lábjegyzete).

FLIGNER és POLICELLO (1981) az FP-próbával végzett szimulációs elemzéseket több folytonos eloszlás (egyenletes, normális, kontaminált normális, dupla exponenciális és Cauchy) felhasználásával. Vizsgált nullhipotézisük a mediánok egyenlősége volt, mely a folytonosság és a szimmetria feltétele esetén ekvivalens a sztochasztikus egyenlőséggel. Eredményeik szerint az FP-próba megtartja érvényességét számottevően különböző szórások esetén is. Az FP-próbával ZUMBO és COULOMBE (1997) is végzett újabb szimulációs elemzést, kis minták generálásával ($m, n \leq 12$). Ők is a mediánok egyenlőségét állították fel nullhipotézisként egyrészt normális, másrészt erősen ferde (exponenciálissal kevert normális) eloszlás felhasználásával. Eredményeik azt mutatták, hogy az FP-próba eléggé inkonzisztens elsőfajú hibákat produkál a ferde eloszlás esetén (vö. ZUMBO, COULOMBE, 1997, 2. táblázat), ami azonban megint csak azzal lehet összefüggésben, hogy különböző elméleti szórások mellett a mediánok egyenlőségét beállítva (alkalmas eltolással), ferde eloszlások esetén rendszerint nem teljesül a sztochasztikus egyenlőség, ami ilyen eloszlásoknál az FP-próba érvényes nullhipotézise (vö. FLIGNER, POLICELLO, 1981, 164). ZUMBO és COULOMBE (1997) arról az érdekes eredményről is beszámol, hogy normális eloszlású függő változók esetén az FP-próba számos olyan esetben is konzervatívan viselkedik (a névlegesnél alacsonyabb elsőfajú hibát szolgáltatva), amikor FLIGNER és POLICELLO (1981) a névlegest jól megközelítő elsőfajú hibaszintet regisztrált. Például $m = 11, n = 10$ és $\alpha = 0,05$ esetén Fligner és Policello elsőfajú hibabecslése 0,048 volt, míg Zumbo és Coulombe ugyanazon feltételek (egyenlő szórás, kétoldali próba) mellett 0,030-as elsőfajú hibabecslést közöl (ZUMBO, COULOMBE, 1997, 1. táblázat).

A szimulációban alapvetően négy különböző rangsorolási csoport-összehasonlító statisztikai próbát használtam a sztochasztikus egyenlőség nullhipotézisének vizsgálatára. A Mann-Whitney-próbával gyakorlatilag egyenértékű rang t-próbát (rt), a rang Welch-próbát (rW), Fligner és Policello próbáját (FP), valamint ez utóbbi Cliff-féle módosított változatát (FPC).

Fligner és Policello felfigyelt arra, hogy tesztstatisztikájuk mennyire hasonlít a Welch-próba tesztstatisztikájára (FLIGNER, POLICELLO, 1981, 164). A fő különbség

a kettő között mindössze abból adódott, hogy a Welch-próba statisztikáját az eredeti megfigyelésekből (X-adatokból) kell kiszámítani, míg az FP-próba statisztikáját a helyértékekből (vö. *Melléklet*). Ők ezt – ugyanúgy, mint ZUMBO és COULOMBE (1997, 140) – elméleti megerősítésként értelmezték az FP-próba különböző szórásokkal szembeni robusztusságával kapcsolatban. A Welch-próba azonban nem csupán képletében különbözik a kétmintás t-próbától, hanem szabadságfokának meghatározásában is. Ez vezet arra a gondolatra, hogy nagyobb minták esetén talán jobb lenne, ha az FP- és az FPC-próbát nem a FLIGNER és POLICELLO (1981), illetve CLIFF (1993) által javasolt normális eloszlás, hanem a Welch-próba esetén is alkalmazott t-eloszlás táblázata alapján értékelnénk ki. A szabadságfok értelemszerűen a helyértékek szórásai felhasználásával, a Welch-próba szabadságfokának ismert képlete (lásd pl. HAJTMAN, 1968, 167; VINCZE, 1968, 131; WILCOX, 1996, 133) alapján határozandó meg (lásd a *Melléklet* (7) formuláját). E megfontolások alapján a szimulációba bevontam az FP- és az FPC-próbának ezt a Welch-próba kiértékeléséhez igazított változatát is, melyeket a továbbiakban FPW-, illetve FPCW-próbának nevezünk. Mindezen statisztikai próbák részletes leírása és képletei cikkünk mellékletében vannak összefoglalva.

2. A SZÁMÍTÓGÉPES SZIMULÁCIÓ MÓDSZERE

A számítógépes programok segítségével végzett szimuláció sokféle jelenség modellezésére alkalmas. Itt most a statisztikai próbák alkalmasságvizsgálatára fogjuk használni. Egy statisztikai próba alkalmasságát, jóságát két fő szempont szerint szokták mérlegelni: I. fajta hibája és ereje megítélésével.

2.1. A statisztikai próbák érvényessége

A statisztikai próbákat úgy szokás megszerkeszteni, hogy rögzített (pl. 5%-os) szignifikanciaszintű döntések során az I. fajta hiba, vagyis a nullhipotézis téves elutasításának valószínűsége egyezzen meg a rögzített szignifikanciaszinttel (azaz pl. 5%-kal). Például a kétmintás t-próba esetében elméletileg is igazolható, hogy ha a függő változó eloszlása normális, az elméleti szórások megegyeznek és az adatmintákat egymástól függetlenül és véletlenszerűen választjuk ki, akkor azonos elméleti átlagok (azaz a nullhipotézis teljesülése) esetén a próba éppen 10%-os valószínűséggel lesz szignifikáns az $\alpha = 0,10$ szinten, 5%-os valószínűséggel lesz szignifikáns az $\alpha = 0,05$ szinten, 1%-os valószínűséggel lesz szignifikáns az $\alpha = 0,01$ szinten stb. Ezt a statisztikai próbákkal szemben megfogalmazott elvárást az *érvényesség* kritériumának nevezzük. Ha egy próba nem érvényes (pl. azért, mert nem tudjuk biztosítani alkalmazási feltételeinek teljesülését), akkor a próba nem a szignifikanciaszint által beállított mértékben (pl. $\alpha = 0,05$ esetén nem az esetek kb. 5%-ában) jelez tévesen szignifikáns különbséget az összevetendő két átlag között, hanem annál gyakrabban vagy ritkábban. Például, ha a kétmintás t-próbával összehasonlított két független minta elemszáma $m = 15$ és $n = 5$, az X függő változó

normális eloszlású és az elméleti átlagok megegyeznek, akkor azonos elméleti szórárok ($\sigma_1 = \sigma_2$) esetén $\alpha = 0,05$ szignifikanciaszint mellett az esetek kb. 5%-ában számíthatunk szignifikáns eredményre, tehát tévesen jelzett elméleti átlagkülönbségre. Ha viszont ugyanezen körülmények mellett az 1. populáció szórása feleakkora, mint a 2. populációé ($\sigma_1 = 0,5\sigma_2$), akkor az I. fajta hiba 5%-ról megnő 7,2%-ra, ha pedig az 1. populáció szórása kétszer akkora, mint a 2. populációé ($\sigma_1 = 2\sigma_2$), akkor az I. fajta hiba 5%-ról lecsökken 3,8%-ra (vö. SCHEFFÉ, 1959, 353, 10.4.1 táblázat).

Egy statisztikai próbát *robosztusnak* nevezünk valamely alkalmazási feltételre vonatkozóan, ha e feltétel nem teljesülése esetén a próba érvényessége nem csökken számottevően. Például a kétmintás t-próba bizonyos mértékig robusztusnak tekinthető a szórás-homogenitás feltételével szemben abban a speciális esetben, ha a mintaelemszámok megegyeznek. Normális eloszlású változóknál ugyanis $m = n = 7$ és 5%-os szignifikanciaszint mellett az I. fajta hiba tényleges szintje $\sigma_1 = 2\sigma_2$ esetén 0,051 (vö. SCHEFFÉ, 1959, 353, 10.4.1 táblázat). A kétmintás t-próba robusztus változataként tartják számon a Welch-féle d-próbát (vö. HAJTMAN, 1968, 166–167; VARGHA, 2000, 236–239), mert ez számos esetben különböző elemszámok és különböző elméleti szórárok mellett is a névlegest jól megközelítő I. fajta hibát produkál. A kétmintás t-próba egyéb robusztus alternatíváival kapcsolatban lásd WILCOX (1996, 8. fejezet).

2.2. A statisztikai próbák ereje

A statisztikai próba érvényessége csak azt garantálja, hogy a szignifikáns eredmények alapján hozott – nullhipotézist elutasító – döntés az esetek többségében (α szignifikanciaszint mellett $1 - \alpha$ valószínűséggel) helyes. Arról viszont, hogy a próba várhatóan szignifikáns lesz-e, ha a nullhipotézis nem igaz, hogy tehát képes-e megbízhatóan kimutatni a nullhipotézis adott esetben hamis voltát, a próba ereje tájékoztat. *A próba ereje* annak a valószínűsége, hogy a próba szignifikáns lesz, ha a nullhipotézis nem igaz. A próba ereje alapvetően függ az elemszám(ok)tól és a nullhipotézistől való eltérés nagyságától, de függ az adott eljárástól, valamint a függő változó eloszlásától is. Például az elméleti átlagok egyenlőségének vizsgálatára egyaránt szokták használni a Welch-féle d-próbát és a trimmelt átlagokkal operáló Yuen-próbát, ha a kétmintás t-próba alkalmazási feltételei nem teljesülnek. E két robusztus próba ereje azonban eltérően függ az eloszlás típusától. Normális eloszlás esetén például, ha $m = n = 25$, $\alpha = 0,05$ és az elméleti átlagok közti különbség pontosan egy szórásnyi, akkor a Welch-próba ereje 93%-os, a Yuen-próbáé pedig valamivel kisebb, 89%-os (20%-os trimmelés mellett). Ha azonban a függő változó eloszlása olyan kontaminált (kevert) normális eloszlású, mely 90%-ban egy μ átlagú és σ szórású, 10%-ban pedig egy ugyancsak μ átlagú, de 10σ szórású normális eloszlásból tevődik össze, akkor a Welch-próba ereje csak 28%-os, míg a Yuen-próbáé még mindig 78%-os (vö. WILCOX, 1996, 150).

A statisztikai próbák szimulációs vizsgálatainak az a lényege, hogy számítógéppel igen sokszor generálva véletlen mintákat, kontrollált feltételek (függő változó

eloszlása, mintanagyság, elemszámok és szórások viszonya stb.) mellett vizsgáljuk empirikusan a tényleges I. fajta hibát és a próba erejét. A jelen szimulációs vizsgálatban a sztochasztikus egyenlőség nullhipotézisére az 1. pontban javasolt hat rangsorolós statisztikai próbát (rt, rW, FP, FPW, FPC, FPCW) az alábbi alponatokban részletezett feltételek variálásával elemeztem.

2.3. Az eloszlás típusa

A szimuláció függő változóinak eloszlását a standardizált lambda eloszláscsaládból választottam ki (vö. RAMBERG, TADIKAMALLA, DUDEWICZ, MYKYTKA, 1979), mely feloleli ferdeségben ($\alpha_3 = \mu_3/\sigma^3$) a 0–2, csúcsosságban ($\alpha_4 = \mu_4/\sigma^4$) pedig az 1,8–15,8 tartományt (μ_3 és μ_4 a harmadik, illetve negyedik centrális momentum). Ez az eloszláscsalád sokféleségével lefedi a gyakorlatban előforduló folytonos eloszlások jelentős hányadát. A lambdaeloszlás különösen alkalmas szimulációs vizsgálatok elvégzésére, mert képlete viszonylag egyszerű, s könnyen programozható (lásd pl. VARGHA, 1996).

A jelen vizsgálathoz a lambdaeloszlások közül három ferdeségi szint (szimmetrikus, mérsékelten aszimmetrikus és erősen aszimmetrikus) mindegyikén három csúcsossági szintet (alacsony, közepes, ill. erős csúcsosság) választottam ki. E kilenc eloszlást meghatározó konkrét ferdeségi és csúcsossági (α_3 , α_4) értékpárokat az 1. táblázat tartalmazza.

1. táblázat. A szimulációban alkalmazott lambdaeloszlások ferdeségi (α_3) és csúcsossági (α_4) értékei

Ferdeség mértéke	Csúcsosság mértéke		
	Alacsony	Közepes	Magas
Szimmetrikus	$\alpha_3 = 0, \alpha_4 = 1,8$	$\alpha_3 = 0, \alpha_4 = 3$	$\alpha_3 = 0, \alpha_4 = 9$
Enyhén aszimmetrikus	$\alpha_3 = 1, \alpha_4 = 3,4$	$\alpha_3 = 1, \alpha_4 = 4,6$	$\alpha_3 = 1, \alpha_4 = 10,6$
Erősen aszimmetrikus	$\alpha_3 = 2, \alpha_4 = 8,6$	$\alpha_3 = 2, \alpha_4 = 9,8$	$\alpha_3 = 2, \alpha_4 = 15,8$

Vegyük észre, hogy az α_4 csúcsossági paraméter lehetséges értéktartománya erősen függ az eloszlás ferdeségétől. Nagyobb ferdeség mellett a csúcsossági értékek lehetséges minimuma és maximuma egyaránt nagyobb, mint alacsonyabb ferdeségi szinten. A közepes csúcsossági szint úgy lett kiválasztva, hogy szimmetrikus esetben a közepes csúcsosság α_4 értéke éppen a normális eloszlásra jellemző $\alpha_4 = 3$ érték legyen. Minden ferdeségi szinten az alacsony és a magas csúcsossági szint a feltüntetett legkisebb és legnagyobb csúcsossági érték RAMBERG és munkatársai (1979) tanulmányának 4. táblázatában, mely a lambdaeloszlások szimulációhoz szükséges paraméterértékeit tartalmazza.

A szimulációban a függő változó eloszlását az egymással összehasonlítandó két mintában egymástól függetlenül variáltam. Így az 1. táblázatban összefoglalt 9 eloszlás a két mintában összesen $9 \times 9 = 81$ eloszláspár-kombinációt eredményezett, amelyeket a szimulációban külön-külön mind megvizsgáltam.

Az 1. táblázatban látható aszimmetrikus eloszlások mind pozitív ferdeségűek ($\alpha_3 > 0$). Annak érdekében, hogy a vizsgált statisztikai próbák alkalmasságát ellentétes ferdeségű eloszlásokra is megvizsgálhassuk, az 1. táblázatban szereplő 6 pozitív ferdeségű eloszlást (2. és 3. sor) összepárosítottam 6 ugyanolyan, de negatív ferdeségű eloszlással minden lehetséges párosításban, mely tehát $6 \times 6 = 36$ újabb megvizsgálandó eloszláspárt eredményezett.

A lambdaeloszlásokat standardizált (elméleti átlag = 0, elméleti szórás = 1) formában generáltam, így a ferdeség előjelének átfordítását egy egyszerű -1 -gyel való beszorzással értem el. Ellentétes ferdeségű eloszláspárok esetén mindig az első tag volt a negatív ferdeségű.

A szimulációban tehát összesen $81 + 36 = 117$ folytonos eloszláspárt vizsgáltam meg.

2.4 Az alkalmazott mintaelemszámok

A mintaelemszámokkal kapcsolatban az átlagos mintanagyságot és a minták egymáshoz viszonyított relatív nagyságát variáltam. Az átlagos mintanagyság $((m + n)/2)$ szerint szerepeltek kis minták $((m + n)/2 = 9)$ és mérsékelt nagyságú minták $((m + n)/2 = 18)$, a relatív mintanagyság szerint pedig egyenlő ($m : n = 1 : 1$) és különböző ($m : n = 1 : 2$) minták. E két szempont együttes figyelembevétele kis minták esetén az $(m = n = 9)$, illetve $(m = 6, n = 12)$, nagy minták esetén pedig az $(m = n = 18)$, illetve $(m = 12, n = 24)$ elemszám-kombinációkat eredményezte. A szimulációt ezzel a 4 elemszám-kombinációval végeztem.

2.5 A szórás heterogenitás mértéke

A két populáció szórásának arányára ($\sigma_1 : \sigma_2$) az alábbi hét arányt alkalmaztam:

$$4 : 1, 3 : 1, 2 : 1, 1 : 1, 1 : 2, 1 : 3, 1 : 4.$$

A legszélsőségesebb esetben tehát a függő változó szórása négyszer akkora az egyik populációban, mint a másikban. Bár ez igen extrém szórás heterogenitási szintet képvisel, a gyakorlat néha ilyeneket is produkál (vö. WILCOX, 1996, 131). A szimulációban a 117 eloszláspárt (vö. 2.3. alpont), a 4 mintaelemszám-párt (vö. 2.4. alpont), s minden ilyen kombinációban mind a 7 szórásarányt bevontam a szimulációba. Ezeket a szórásarányokat úgy értem el, hogy a szimuláció során standardizált (tehát 1 szórású) lambdaeloszlású változókat rendre beszoroztam az aránypárban szereplő megfelelő értékekkel.

2.6. A sztochasztikus egyenlőség beállítása

Egy bizonyos összehasonlításhoz először a két mintának megfelelő két eloszlást kellett meghatározni (vö. 2.3. alpont). Ezek mind standardizált folytonos lambda-eloszlások voltak. Ellenkező ferdeségű párokat az első mintához tartozó eloszlás -1 -gyel való beszorzásával képeztem, ami a standardizált formát megőrizte, tehát az eloszlás továbbra is 0 elméleti átlagú volt. A szórások beállítása az eloszlások különböző állandókkal való beszorzásával történt (vö. 2.5. alpont), ami a 0 átlagon ugyancsak nem módosított.

Ha az összevetendő két eloszlás szimmetrikus, akkor az elméleti átlagok egyenlősége egyben a sztochasztikus egyenlőséget is biztosítja. Ellenkező esetben viszont a sztochasztikus egyenlőség nem feltétlenül áll fenn, így ezt a szimuláció előtt be kell állítani. Ezt a beállítást úgy végeztem el, hogy a két eloszlás közül a 2. mintához tartozót eltoltam olyan mértékben, hogy a sztochasztikus egyenlőség fennálljon. Az eltolást úgy valósítottam meg, hogy a 2. minta véletlen változójához hozzáadtam egy alkalmas – pozitív vagy negatív – számot, az ún. *eltolási állandót*, mely persze minden eloszláspár esetén más és más volt. Az eltolási állandókat külön – Turbo Pascal nyelven írt – számítógépes programmal határoztam meg. Ebben egy kezdeti 0 értéket módosítottam iterációs algoritmussal több lépésben úgy, hogy a valószínűségi főlény A_{12} mutatójának pontbecslése minél jobban meg-

2. táblázat. A sztochasztikus egyenlőséget biztosító eltolási állandók néhány eloszláspárra

Eloszlástípus		$\sigma_1 : \sigma_2$ szórásarány						
1. minta	2. minta	4 : 1	3 : 1	2 : 1	1 : 1	1 : 2	1 : 3	1 : 4
$\alpha_3 = 0$ $\alpha_4 = 3$	$\alpha_3 = 0$ $\alpha_4 = 1,8$	0	0	0	0	0	0	0
$\alpha_3 = 0$ $\alpha_4 = 9$	$\alpha_3 = 1$ $\alpha_4 = 10,6$	0,01	0,02	0,03	0,05	0,12	0,21	0,28
$\alpha_3 = 1$ $\alpha_4 = 4,6$	$\alpha_3 = 1$ $\alpha_4 = 4,6$	-0,62	-0,43	-0,23	0	0,23	0,43	0,62
$\alpha_3 = -1$ $\alpha_4 = 4,6$	$\alpha_3 = 1$ $\alpha_4 = 4,6$	0,64	0,46	0,30	0,17	0,30	0,47	0,64
$\alpha_3 = 1$ $\alpha_4 = 4,6$	$\alpha_3 = 2$ $\alpha_4 = 8,6$	-0,61	-0,42	-0,22	0,04	0,41	0,77	1,11
$\alpha_3 = -1$ $\alpha_4 = 4,6$	$\alpha_3 = 2$ $\alpha_4 = 8,6$	0,65	0,49	0,34	0,24	0,49	0,78	1,11
$\alpha_3 = 2$ $\alpha_4 = 8,6$	$\alpha_3 = 2$ $\alpha_4 = 15,8$	-1,11	-0,78	-0,43	-0,07	0,18	0,38	0,58
$\alpha_3 = -2$ $\alpha_4 = 8,6$	$\alpha_3 = 2$ $\alpha_4 = 15,8$	1,13	0,83	0,52	0,26	0,35	0,49	0,64

közelítse a 0,5-öt. Egy ilyen állandót akkor fogadtam el, ha a pontbecslés három egymást követő független becslés során legfeljebb 0,001-gyel különbözött 0,5-től, e végső fázisban iterációs lépésenként 400 000 független adatpárt generálva a beállított eloszláspárból.

Illusztráció céljából néhány eloszláspárhoz tartozó eltolási állandót (két tizedes jegyre kerekítve) a 2. táblázatban foglaltunk össze. Itt vegyük észre, hogy az eltolási állandó csak akkor 0, ha mindkét eloszlás szimmetrikus (lásd első sor) vagy ha mindkettő ferde, de egymással teljesen megegyező eloszlású, amelyeknek a szórásai is megegyeznek (lásd 3. sor 1 : 1 szórásarányhoz tartozó rovata).

2.7. A sztochasztikus egyenlőtlenség különböző szintjeinek a beállítása

A próbák I. fajta hibájának a becsléséhez sztochasztikusan egyenlő eloszlásokat kell biztosítani (ahol is $A_{12} = 0,5$), az erő becsléséhez azonban sztochasztikusan különböző eloszlásokat (ahol is $A_{12} \neq 0,5$). Tekintve, hogy a szimulációs vizsgálatban a pszichológiai kutatásokra igencsak jellemző kicsi és közepes nagyságú mintákat vontam be, a sztochasztikus egyenlőséggel szembeni alternatívaként egy közepes és egy erős sztochasztikus egyenlőtlenségi szintet állítottam be, az $A_{12} = 0,64$ és az $A_{12} = 0,71$ valószínűségi fölény értékek választásával. Normális eloszlások esetén (amikor az elméleti átlagok egyenlősége ekvivalens a sztochasztikus egyenlőséggel) e két szint pontosan megfelel az átlagok Cohen-féle Δ hatásmértékkel mért közepes ($\Delta = 0,5$) és erős ($\Delta = 0,8$) különbségének (vö. VARGHA, 1999a, 1. táblázat). E sztochasztikus egyenlőtlenségi szinteknek megfelelő eloszláspárok szintén alkalmas eltolási állandókkal készíthetők, amelyeket az $A_{12} = 0,5$ esettel analóg módon határoztam meg (vö. 2.6. alpont).

2.8. A szimuláció végrehajtása

A 117 különböző eloszláspár (vö. 2.3. alpont) mindegyikét a 7 szórás-kombináció (vö. 2.5. alpont) mindegyikével párosítva 4 mintaelemszám kombinációban (vö. 2.4. alpont) vizsgáltam meg egyrészt a sztochasztikus egyenlőség nullhipotézise, vagyis $A_{12} = 0,5$ igaz volta (vö. 2.6. alpont), másrészt a sztochasztikus egyenlőtlenség két szintje ($A_{12} = 0,64$ és $A_{12} = 0,71$) mellett.

Minden egyes ilyen kombinációban (ezek száma összesen $117 \times 7 \times 4 \times 3 = 9828$ volt) egymástól függetlenül igen sokszor generáltam két független adatmintát (a kombinációnak megfelelő eloszlásokkal, szórásarányal, elemszámokkal és A_{12} értékkel), kiszámítottam a hat vizsgált statisztikai eljárás próbastatisztikáját (vö. Melléklet), s megnéztem, hogy szignifikánsak-e 5, illetve 10%-os valószínűségi szinten, kétoldalú ellenhipotézist alkalmazva. Az r_t -, r_W -, FPW-, FPCW-próbák statisztikáját a t-táblázat megfelelő szabadságfokhoz tartozó táblázati értéke alapján értékeltem ki. FP és FPC esetében, kis minták ($m + n = 18$) mellett Fligner és Policello e célra készített táblázatát (lásd FLIGNER, POLICELLO, 1981, 1. táblázat),

közepes minták ($m + n = 36$) esetén pedig a standard normális eloszlás táblázatát alkalmaztam (vö. WILCOX, 1996, 370).

Végül meghatároztam minden kombinációban mind a két valószínűségi szinten a szignifikáns eredmények arányát. Ezek az arányok $A_{12} = 0,5$ mellett az egyes próbák I. fajta hibáját, $A_{12} \neq 0,5$ mellett pedig a próbák erejét becsülik. A független ismétlések száma az I. fajta hiba vizsgálata során 100 000 volt, ami mellett a becslés átlagos hibája sem 5, sem 10%-os szinten nem éri el a 0,001-et (vö. VARGHA, 1996). Az I. fajta hibabecslések tehát a valódi (elméleti) értékektől várhatóan csak a 3. tizedes jegyben különböznek, s az eltérés ott is csak legfeljebb 1-2 jegy (azaz 0,001–0,002). Az ismétlési szám az erővizsgálatok során 20 000 volt, ami mellett a becslés átlagos hibája sem 5, sem 10%-os szinten nem haladja meg a 0,005-öt (vö. VARGHA, 1996). Az erőbecslések tehát a valódi (elméleti) értékektől várhatóan csak a 3. tizedes jegyben különböznek. A szimulációs elemzéseket egy 200 MHz-es Pentium PC-n végeztem el. A lambdaeloszlásokat a RAMBERG és munkatársai (1979) tanulmányában leírt módon generáltam, s az ehhez szükséges pseudo-véletlen egyenletes eloszlást a Turbo Pascalba beépített Random függvénnyel hoztam létre. Ez lineáris kongruens véletlenszám-generátor, mely egy közelmúltban publikált tanulmány szerint több vonatkozásban a legjobb eljárások egyikének tekinthető (vö. ONGHENA, 1993).

3. EREDMÉNYEK

Egy statisztikai próba alkalmasságát érvényességének és hatékonyságának együttes figyelembevételével lehet megítélni, így az I. fajta hibára és az erőre vonatkozó eredményeket együtt fogjuk kiértékelni. Először az azonos nagyságú minták esetét tárgyaljuk (lásd 3.1. alpont), majd a különböző nagyságú mintákra vonatkozó eredményeket ismertetjük (lásd 3.2. alpont). Ez utóbbi azért igényel külön elemzést, mert különböző mintaelemszámok esetén a mintaelemszámok és szórások viszonya az egyik legalapvetőbb meghatározója a kétmintás t-próba I. fajta hibájának (vö. SCHEFFÉ, 1959, 353, 10.4.1 táblázat).

3.1. Egyenlő mintaelemszámok esete

A szimulációs vizsgálatba összesen 117 különböző folytonos eloszláspárt vontam be, szisztematikusan variálva a ferdeség és a csúcosság mértékét és ezek egymáshoz viszonyított arányát a két mintában (vö. 2.3. alpont). Az eredmények érdekes módon azt mutatták, hogy – r_t -vel ellentétben – a robusztus próbák I. fajta hibáját nem befolyásolja számottevően a függő változó eloszlása a két minta által képviselt populációban. Ez okból az eredmények könnyebb áttekintése érdekében az I. fajta hibákat összesítettem a 117 eloszláspárra, kiszámítva átlagukat, valamint legkisebb és legnagyobb értéküket. Ezeket az összesítéseket a hat vizsgált próbára az alkalmazott szórásarányok szerinti bontásban kis minták ($m = n = 9$) esetén a 3. táblázat, közepes minták ($m = n = 18$) esetén pedig a 4. táblázat tartalmazza. Tekintve,

hogyan azonos nagyságú minták esetén az 1:2 és a 2:1, az 1:3 és a 3:1, illetve az 1:4 és a 4:1 szórásarány a szórások különbözőségének egymással teljesen egyenértékű mértéke, az ezekhez tartozó eredményeket páronként összevontam, így különböző szórások esetén a 3. és a 4. táblázatban összefoglalt eredmények $2 \times 117 = 234$ különböző eloszláspárra vonatkozó összesítések. Az ugyanezen elrendezésekben kapott erőbecslések átlagát (a 117, illetve 234 eloszláspárra vonatkozóan) a $H_1: A_{12} = 0,64$ ellenhipotézis (vagyis a nullhipotézistől való közepes mértékű eltérés beállítása) mellett az 5. táblázat, a $H_1: A_{12} = 0,71$ ellenhipotézis (vagyis a nullhipotézistől való erős eltérés beállítása) mellett a 6. táblázat tartalmazza. Mindezen táblázatok alapján a következő megállapításokat tehetjük.

3. táblázat. A vizsgált 6 próba I. fajta hibája kicsi, egyenlő mintaelemszámok esetén ($m = n = 9$), 117 (eltérő szórások esetén 234) különböző eloszláspárra összesítve

I. fajta hibabecslés átlagok ($\alpha = 0,05$)					I. fajta hibabecslés átlagok ($\alpha = 0,10$)				
$(\sigma_1 : \sigma_2)$	(1 : 1)	(1 : 2)	(1 : 3)	(1 : 4)	$(\sigma_1 : \sigma_2)$	(1 : 1)	(1 : 2)	(1 : 3)	(1 : 4)
rt:	0,052	0,060	0,069+	0,077++	rt:	0,097	0,106	0,118	0,127+
rW:	0,052	0,058	0,063+	0,066+	rW:	0,097	0,106	0,118	0,127+
FP:	0,051	0,054	0,057	0,057	FP:	0,101	0,103	0,105	0,108
FPW:	0,054	0,055	0,056	0,055	FPW:	0,097	0,097	0,098	0,099
FPC:	0,059	0,061+	0,062+	0,062+	FPC:	0,113	0,116	0,119	0,123+
FPCW:	0,063+	0,062+	0,060	0,058	FPCW:	0,109	0,107	0,105	0,104

I. fajta hibabecslés minimumok ($\alpha = 0,05$)					I. fajta hibabecslés minimumok ($\alpha = 0,10$)				
$(\sigma_1 : \sigma_2)$	(1 : 1)	(1 : 2)	(1 : 3)	(1 : 4)	$(\sigma_1 : \sigma_2)$	(1 : 1)	(1 : 2)	(1 : 3)	(1 : 4)
rt:	0,050	0,052	0,059	0,065	rt:	0,092	0,096	0,105	0,115
rW:	0,049	0,052	0,057	0,061	rW:	0,092	0,096	0,105	0,115
FP:	0,050	0,051	0,054	0,054	FP:	0,099	0,099	0,101	0,102
FPW:	0,053	0,053	0,053	0,051	FPW:	0,095	0,095	0,096	0,096
FPC:	0,058	0,059	0,059	0,058	FPC:	0,111	0,112	0,114	0,116
FPCW:	0,061	0,058	0,055	0,052	FPCW:	0,106	0,102	0,102	0,101

I. fajta hibabecslés maximumok ($\alpha = 0,05$)					I. fajta hibabecslés maximumok ($\alpha = 0,10$)				
$(\sigma_1 : \sigma_2)$	(1 : 1)	(1 : 2)	(1 : 3)	(1 : 4)	$(\sigma_1 : \sigma_2)$	(1 : 1)	(1 : 2)	(1 : 3)	(1 : 4)
rt:	0,057	0,074	0,086	0,096	rt:	0,103	0,122	0,136	0,145
rW:	0,056	0,067	0,069	0,069	rW:	0,103	0,122	0,136	0,145
FP:	0,054	0,059	0,059	0,060	FP:	0,104	0,108	0,113	0,119
FPW:	0,056	0,058	0,058	0,058	FPW:	0,101	0,100	0,102	0,106
FPC:	0,061	0,064	0,064	0,065	FPC:	0,117	0,121	0,128	0,135
FPCW:	0,065	0,064	0,063	0,062	FPCW:	0,113	0,111	0,109	0,108

Megjegyzés: A táblázatban a 20, illetve 40%-os mértéket meghaladó I. fajta hibabecslésátlagot +, illetve ++ jelöli.

4. táblázat. A vizsgált 6 próba I. fajta hibája közepes nagyságú, egyenlő mintaelemszámok esetén ($m = n = 18$), 117 (eltérő szórások esetén 234) különböző eloszláspárra összesítve

I. fajta hibabecslés átlagok ($\alpha = 0,05$)					I. fajta hibabecslés átlagok ($\alpha = 0,10$)				
$(\sigma_1 : \sigma_2)$	(1 : 1)	(1 : 2)	(1 : 3)	(1 : 4)	$(\sigma_1 : \sigma_2)$	(1 : 1)	(1 : 2)	(1 : 3)	(1 : 4)
rt:	0,053	0,060	0,069+	0,076++	rt:	0,100	0,110	0,122+	0,131+
rW:	0,053	0,059	0,067+	0,072++	rW:	0,100	0,110	0,122+	0,131+
FP:	0,060	0,061+	0,063+	0,064+	FP:	0,107	0,109	0,111	0,113
FPW:	0,051	0,051	0,051	0,051	FPW:	0,097	0,097	0,097	0,098
FPC:	0,065+	0,066+	0,067+	0,067+	FPC:	0,114	0,115	0,116	0,116
FPCW:	0,056	0,055	0,054	0,054	FPCW:	0,104	0,103	0,102	0,101

I. fajta hibabecslés minimumok ($\alpha = 0,05$)					I. fajta hibabecslés minimumok ($\alpha = 0,10$)				
$(\sigma_1 : \sigma_2)$	(1 : 1)	(1 : 2)	(1 : 3)	(1 : 4)	$(\sigma_1 : \sigma_2)$	(1 : 1)	(1 : 2)	(1 : 3)	(1 : 4)
rt:	0,050	0,053	0,060	0,067	rt:	0,095	0,099	0,110	0,120
rW:	0,050	0,053	0,059	0,065	rW:	0,095	0,099	0,110	0,120
FP:	0,058	0,059	0,061	0,062	FP:	0,103	0,106	0,108	0,110
FPW:	0,049	0,049	0,049	0,049	FPW:	0,094	0,095	0,095	0,095
FPC:	0,063	0,064	0,065	0,065	FPC:	0,111	0,112	0,113	0,114
FPCW:	0,054	0,053	0,052	0,051	FPCW:	0,102	0,100	0,099	0,099

I. fajta hibabecslés maximumok ($\alpha = 0,05$)					I. fajta hibabecslés maximumok ($\alpha = 0,10$)				
$(\sigma_1 : \sigma_2)$	(1 : 1)	(1 : 2)	(1 : 3)	(1 : 4)	$(\sigma_1 : \sigma_2)$	(1 : 1)	(1 : 2)	(1 : 3)	(1 : 4)
rt:	0,059	0,073	0,083	0,088	rt:	0,108	0,127	0,139	0,147
rW:	0,059	0,071	0,078	0,083	rW:	0,108	0,127	0,139	0,147
FP:	0,062	0,065	0,067	0,068	FP:	0,110	0,114	0,116	0,117
FPW:	0,053	0,053	0,053	0,053	FPW:	0,100	0,100	0,100	0,101
FPC:	0,067	0,069	0,069	0,070	FPC:	0,117	0,118	0,119	0,119
FPCW:	0,058	0,057	0,056	0,056	FPCW:	0,107	0,106	0,104	0,104

Megjegyzés: A táblázatban a névleges szintet több mint 20, illetve 40%-os mértéket meghaladó I. fajta hibabecslésátlagot +, illetve ++ jelöli.

1. Kicsi és ugyanakkora minták ($m = n = 9$) esetén 5%-os szignifikanciaszint mellett csak két próbára, FP-re és FPW-re teljesül az, hogy I. fajta hibaszintje még a legszélsőségesebb szórásarány mellett sem tér el 20%-osnál nagyobb mértékben a névleges szinttől, vagyis hogy mindig 0,04 és 0,06 közé esik. A minimális és maximális becslések azt mutatják, hogy FP és FPW esetében ez nem csak az átlagra, hanem az összes megvizsgált eloszláspárra külön-külön is igaz. Az FPW-próba csekély mértékben jobbnak tűnik FP-nél, mivel esetében az I. fajta hibák 0,051 és 0,058 között ingadozva kevésbé térnek el a névleges szinttől, mint az FP-próba hibái. Ez utóbbiak ugyanis 0,050 és 0,060 között ingadoznak és a szórásarány élesebbé válásával határozott növekvő tendenciát mutatnak (vö. 3. táblázat, $\alpha = 5\%$ melletti hibabecslésátlagok). Mivel FP és FPW ereje $m = n = 9$ esetén gyakorlatilag

5. táblázat. A vizsgált 6 próba ereje egyenlő mintaelemszámok mellett, 117 (eltérő szórások esetén 234) különböző eloszláspárra leátlagolva a $H_1 : A_{12} = 0,64$ ellenhipotézis igaz volta esetén

Erőátlagok $\alpha = 0,05$ és $m = n = 9$ mellett					Erőátlagok $\alpha = 0,05$ és $m = n = 18$ mellett				
$(\sigma_1 : \sigma_2)$	(1 : 1)	(1 : 2)	(1 : 3)	(1 : 4)	$(\sigma_1 : \sigma_2)$	(1 : 1)	(1 : 2)	(1 : 3)	(1 : 4)
rt:	0,17	0,18	(0,19)	(0,20)	rt:	0,31	0,31	(0,32)	(0,32)
rW:	0,17	0,18	(0,18)	(0,18)	rW:	0,31	0,31	(0,31)	(0,31)
FP:	0,17	0,17	0,16	0,16	FP:	(0,32)	(0,31)	(0,30)	(0,29)
FPW:	0,17	0,17	0,16	0,15	FPW:	0,30	0,28	0,27	0,26
FPC:	0,18	(0,18)	(0,17)	(0,17)	FPC:	(0,34)	(0,32)	(0,31)	(0,30)
FPCW:	(0,19)	(0,18)	0,17	0,16	FPCW:	0,31	0,29	0,28	0,26

Erőátlagok $\alpha = 0,10$ és $m = n = 9$ mellett					Erőátlagok $\alpha = 0,10$ és $m = n = 18$ mellett				
$(\sigma_1 : \sigma_2)$	(1 : 1)	(1 : 2)	(1 : 3)	(1 : 4)	$(\sigma_1 : \sigma_2)$	(1 : 1)	(1 : 2)	(1 : 3)	(1 : 4)
rt:	0,26	0,26	0,27	(0,28)	rt:	0,42	0,42	(0,42)	(0,43)
rW:	0,26	0,26	0,27	(0,28)	rW:	0,42	0,42	(0,42)	(0,43)
FP:	0,26	0,25	0,25	0,25	FP:	0,43	0,42	0,40	0,39
FPW:	0,25	0,25	0,24	0,23	FPW:	0,41	0,39	0,38	0,36
FPC:	0,28	0,28	0,27	(0,27)	FPC:	0,44	0,43	0,41	0,40
FPCW:	0,27	0,26	0,25	0,24	FPCW:	0,42	0,40	0,38	0,37

6. táblázat. A vizsgált 6 próba ereje egyenlő mintaelemszámok mellett, 117 (eltérő szórások esetén 234) különböző eloszláspárra leátlagolva a $H_1 : A_{12} = 0,71$ ellenhipotézis igaz volta esetén

Erőátlagok $\alpha = 0,05$ és $m = n = 9$ mellett					Erőátlagok $\alpha = 0,05$ és $m = n = 18$ mellett				
$(\sigma_1 : \sigma_2)$	(1 : 1)	(1 : 2)	(1 : 3)	(1 : 4)	$(\sigma_1 : \sigma_2)$	(1 : 1)	(1 : 2)	(1 : 3)	(1 : 4)
rt:	0,33	0,34	(0,35)	(0,36)	rt:	0,60	0,60	(0,59)	(0,59)
rW:	0,33	0,33	(0,33)	(0,32)	rW:	0,60	0,60	(0,59)	(0,58)
FP:	0,32	0,32	0,31	0,30	FP:	(0,62)	(0,60)	(0,57)	(0,56)
FPW:	0,33	0,32	0,30	0,29	FPW:	0,59	0,56	0,53	0,51
FPC:	0,35	(0,34)	(0,32)	(0,31)	FPC:	(0,63)	(0,61)	(0,58)	(0,56)
FPCW:	(0,35)	(0,33)	0,31	0,29	FPCW:	0,60	0,57	0,54	0,52

Erőátlagok $\alpha = 0,10$ és $m = n = 9$ mellett					Erőátlagok $\alpha = 0,10$ és $m = n = 18$ mellett				
$(\sigma_1 : \sigma_2)$	(1 : 1)	(1 : 2)	(1 : 3)	(1 : 4)	$(\sigma_1 : \sigma_2)$	(1 : 1)	(1 : 2)	(1 : 3)	(1 : 4)
rt:	0,45	(0,45)	0,45	(0,46)	rt:	0,71	0,71	(0,70)	(0,70)
rW:	0,45	0,45	0,45	(0,46)	rW:	0,71	0,71	(0,70)	(0,70)
FP:	0,45	0,44	0,42	0,42	FP:	0,72	0,70	0,68	0,66
FPW:	0,44	0,42	0,41	0,40	FPW:	0,70	0,68	0,65	0,64
FPC:	0,47	(0,46)	0,45	(0,45)	FPC:	0,73	0,71	0,68	0,67
FPCW:	(0,46)	(0,44)	0,42	0,41	FPCW:	0,71	0,69	0,66	0,64

Megjegyzés: Mindkét táblázatban zárójelbe tettünk minden olyan erőátlagot, amelyhez tartozó megfelelő I. fajta hibaátlag (vö. 3. és 4. táblázat) több mint 20%-kal meghaladja a névleges szintet.

ugyanolyan szintű (vö. 5. és 6. táblázat bal felső blokkja), kicsi és egyenlő minták esetén az FPW-próba tűnik a legjobbnak.

2. Ugyanezen elrendezésekben $\alpha = 10\%$ esetén FP és FPW mellett az FPCW-próba I. fajta hibabecslésátlaga marad még minden szórásarány esetén 20%-os eltérésen belül (azaz 0,080 és 0,120 között). Közülük azonban megint FPW tűnik a pontosabbnak, melynek I. fajta hibái – 0,095 és 0,106 között ingadozva – egy esetben sem térnek el 6%-osnál jobban a névleges szinttől. A maximális eltérés FP esetében 19%, FPCW esetében pedig 13% (lásd 3. táblázat jobb oldala). Mivel pedig e három próba ereje 10%-os szignifikanciaszint mellett ismét azonos nagyságrendű (vö. 5. és 6. táblázat bal alsó blokkja), kicsi és egyenlő minták esetén 10%-os szinten is az FPW-próba ajánlható leginkább a sztochasztikus egyenlőség nullhipotézisének vizsgálatára.

3. Közepes nagyságú és ugyanakkora minták ($m = n = 18$) esetén, 5%-os szignifikanciaszinten csak az FPW- és az FPCW-próbára teljesül az, hogy I. fajta hibaszintje még a legszélsőségesebb szórásarány mellett sem tér el 20%-osnál nagyobb mértékben a névleges szinttől. Előbbi I. fajta hibái 0,049 és 0,053, utóbbié pedig 0,051 és 0,058 között ingadoznak (vö. 4. táblázat bal oldala), s mivel erejük teljesen egy szinten mozog (vö. 5. és 6. táblázat jobb felső blokkja), a legjobb próba itt is FPW, bár FPCW alig marad el tőle.

4. Ugyanezen elrendezésekben, de $\alpha = 10\%$ esetén négy próba (FP, FPW, FPC, FPCW) I. fajta hibái maradnak minden szórásarány esetén 20%-os eltérésen belül. Közülük azonban kiemelkedik FPW és FPCW, amelyek esetében a maximális eltérés rendre 5, illetve 7%-os (vö. 4. táblázat jobb oldala). Az erő tekintetében azt tapasztalhatjuk, hogy FP és FPC ereje általában 2-3 százalékponttal nagyobb, mint a másik kettőé (vö. 5. és 6. táblázat jobb alsó blokkja), de ez kizárólag az I. fajta hibájuk megemelkedett szintjének köszönhető. Ez okból közepes nagyságú egyenlő minták esetén 10%-os szinten is az FPW- és az FPCW-próba ajánlható leginkább.

A fenti eredmények fényében egyértelműnek tűnik, hogy a sztochasztikus egyenlőség nullhipotézisének vizsgálatára azonos mintaelemszámok esetén az FPW-próba a legalkalmasabb, melynek az I. fajta hibája valamennyi vizsgált eloszlás, illetve elemszám-szórás kombináció mellett legfeljebb 16%-kal tér el a névleges szinttől. Imponáló, hogy FPW ingadozási sávja $\alpha = 5\%$ mellett (0,049-0,058), $\alpha = 10\%$ mellett (0,095-0,106), ereje pedig a többi még elfogadható érvényességű eljárással azonos szintű.

Az FP-próba FPW-hez hasonló jóságú kis minták esetén, csak a szórás heterogenitásra érzékenyebb némileg jobban, mint FPW. Közepes minták esetén ugyanakkor FP I. fajta hibája érezhetően megnő, s emiatt itt már nem versenyképes FPW-vel szemben.

Az FPCW-próba némileg ellentétesen viselkedik, mint FP. Kis minták esetén számos esetben magasabb az I. fajta hibaszintje, mint FPW-nek, a mintanagyság megemelkedésével azonban FPW-hez hasonló teljesítményt mutat.

Az FPC-próba nem produkál a névleges szinttől drasztikusan eltérő I. fajta hibákat (az eltérés sehol sem haladja meg a 40%-os mértéket, így pl. $\alpha = 5\%$ mellett a 0,07-et), de ez a teljesítmény határozottan gyengébb, mint a másik háromé.

Az rW-próba összesítésben hasonló teljesítményű, mint FPC. Előnyösebb I. fajta hibaszint jellemzi azonos szórások mellett, viszont erős szórás heterogenitás

esetén az I. fajta hibája jobban megugrik, mint FPC-é, tehát nem annyira robusztus a szórások különbözőségével szemben.

Az rt-próba a kétmintás t-próbához hasonlóan teljesen elfogadható eljárás azonos elméleti szórások mellett, ellenkező esetben viszont az I. fajta hibája – esetenként már 1:2 szórásarány mellett is – jelentősen megemelkedik.

3.2. Különböző mintaelemszámok esete

Az eredmények itt is azt mutatták, hogy a robusztus próbák I. fajta hibáját nem befolyásolja lényeges mértékben a függő változó eloszlása. Ez esetben az összesített I. fajta hibaszinteket a hat vizsgált próbára az alkalmazott 7 szórásarány szerinti bontásban kis minták ($m = 6, n = 12$) esetén a 7. táblázat, közepes minták ($m = 12, n = 24$) esetén pedig a 8. táblázat tartalmazza. Az ugyanezen elrendezésekben kapott erőbecslések átlagát (a 117 eloszláspárra vonatkozóan) a $H_1: A_{12} = 0,64$ ellenhipotézis (vagyis a nullhipotézistől való közepes mértékű eltérés beállítása) mellett a 9. táblázat, a $H_1: A_{12} = 0,71$ ellenhipotézis (vagyis a nullhipotézistől való erős eltérés beállítása) mellett a 10. táblázat tartalmazza.

7. táblázat. A vizsgált 6 próba I. fajta hibája kicsi, különböző mintaelemszámok esetén ($m=6, n=12$), 117 különböző eloszláspárra összesítve

I. fajta hibabecslés átlagok ($\alpha = 0,05$)							
$(\sigma_1:\sigma_2)$	(4:1)	(3:1)	(2:1)	(1:1)	(1:2)	(1:3)	(1:4)
rt:	0,131++	0,114++	0,090++	0,055	0,037	0,034	0,034
rW:	0,063+	0,064+	0,064+	0,059	0,061+	0,069+	0,077++
FP:	0,064+	0,066+	0,065+	0,052	0,040	0,037	0,037
FPW:	0,056	0,058	0,058	0,051	0,046	0,047	0,048
FPC:	0,061+	0,065+	0,066+	0,056	0,045	0,042	0,042
FPCW:	0,057	0,061+	0,063+	0,060	0,055	0,054	0,054
I. fajta hibabecslés minimumok ($\alpha = 0,05$)							
$(\sigma_1:\sigma_2)$	(4:1)	(3:1)	(2:1)	(1:1)	(1:2)	(1:3)	(1:4)
rt:	0,112	0,093	0,070	0,041	0,033	0,032	0,032
rW:	0,057	0,060	0,061	0,055	0,054	0,057	0,065
FP:	0,058	0,061	0,059	0,042	0,036	0,035	0,035
FPW:	0,049	0,053	0,056	0,047	0,043	0,043	0,045
FPC:	0,052	0,057	0,063	0,048	0,041	0,040	0,040
FPCW:	0,049	0,053	0,058	0,056	0,052	0,051	0,051
I. fajta hibabecslés maximumok ($\alpha = 0,05$)							
$(\sigma_1:\sigma_2)$	(4:1)	(3:1)	(2:1)	(1:1)	(1:2)	(1:3)	(1:4)
rt:	0,160	0,144	0,120	0,078	0,045	0,038	0,038
rW:	0,068	0,069	0,069	0,066	0,073	0,085	0,093
FP:	0,069	0,070	0,069	0,062	0,045	0,041	0,040
FPW:	0,062	0,063	0,062	0,059	0,051	0,051	0,050
FPC:	0,070	0,072	0,070	0,065	0,051	0,047	0,044
FPCW:	0,065	0,068	0,068	0,064	0,059	0,057	0,056

I. fajta hibabecslés átlagok ($\alpha = 0,10$)							
$(\sigma_1:\sigma_2)$	(4:1)	(3:1)	(2:1)	(1:1)	(1:2)	(1:3)	(1:4)
rt:	0,192++	0,177++	0,151++	0,105	0,079	0,075	0,076
rW:	0,127+	0,118	0,110	0,104	0,111	0,122+	0,130+
FP:	0,142++	0,131+	0,119	0,101	0,087	0,084	0,085
FPW:	0,112	0,105	0,100	0,092	0,086	0,086	0,087
FPC:	0,132+	0,125+	0,119	0,110	0,100	0,096	0,095
FPCW:	0,081	0,089	0,098	0,102	0,102	0,103	0,105

I. fajta hibabecslés minimumok ($\alpha = 0,10$)							
$(\sigma_1:\sigma_2)$	(4:1)	(3:1)	(2:1)	(1:1)	(1:2)	(1:3)	(1:4)
rt:	0,177	0,158	0,126	0,085	0,072	0,071	0,071
rW:	0,116	0,110	0,105	0,101	0,100	0,107	0,117
FP:	0,130	0,122	0,110	0,091	0,082	0,081	0,081
FPW:	0,104	0,100	0,096	0,087	0,082	0,082	0,083
FPC:	0,122	0,118	0,114	0,102	0,095	0,092	0,092
FPCW:	0,067	0,075	0,087	0,097	0,095	0,097	0,100

I. fajta hibabecslés maximumok ($\alpha = 0,10$)							
$(\sigma_1:\sigma_2)$	(4:1)	(3:1)	(2:1)	(1:1)	(1:2)	(1:3)	(1:4)
rt:	0,211	0,202	0,183	0,137	0,090	0,081	0,084
rW:	0,149	0,136	0,121	0,110	0,126	0,136	0,141
FP:	0,164	0,152	0,135	0,113	0,095	0,089	0,089
FPW:	0,132	0,120	0,107	0,099	0,092	0,091	0,092
FPC:	0,151	0,140	0,127	0,117	0,106	0,101	0,098
FPCW:	0,090	0,099	0,107	0,108	0,106	0,108	0,110

Megjegyzés: A táblázatban a névleges szintet több mint 20, illetve 40%-os mértékkel meghaladó I. fajta hibabecslés átlagot +, illetve ++ jelöli.

8. táblázat. A vizsgált 6 próba I. fajta hibája közepes nagyságú, különböző mintaelemszámok esetén ($m=12, n=24$), 117 különböző eloszláspárra összesítve

I. fajta hibabecslés átlagok ($\alpha = 0,05$)							
$(\sigma_1:\sigma_2)$	(4:1)	(3:1)	(2:1)	(1:1)	(1:2)	(1:3)	(1:4)
rt:	0,116++	0,103++	0,083++	0,051	0,034	0,032	0,032
rW:	0,066+	0,063+	0,058	0,054	0,060	0,071++	0,080++
FP:	0,078++	0,077++	0,074++	0,065+	0,058	0,056	0,056
FPW:	0,055	0,054	0,052	0,048	0,046	0,047	0,047
FPC:	0,076++	0,076++	0,075++	0,070+	0,065+	0,063+	0,063+
FPCW:	0,054	0,053	0,053	0,052	0,052	0,053	0,054

I. fajta hibabecslés minimumok ($\alpha = 0,05$)							
$(\sigma_1:\sigma_2)$	(4:1)	(3:1)	(2:1)	(1:1)	(1:2)	(1:3)	(1:4)
rt:	0,103	0,087	0,065	0,037	0,030	0,029	0,030
rW:	0,062	0,059	0,054	0,050	0,053	0,060	0,069
FP:	0,075	0,074	0,070	0,060	0,055	0,054	0,054
FPW:	0,053	0,052	0,049	0,045	0,043	0,044	0,045
FPC:	0,072	0,074	0,072	0,067	0,062	0,062	0,061
FPCW:	0,051	0,052	0,051	0,050	0,049	0,051	0,052
I. fajta hibabecslés maximumok ($\alpha = 0,05$)							
$(\sigma_1:\sigma_2)$	(4:1)	(3:1)	(2:1)	(1:1)	(1:2)	(1:3)	(1:4)
rt:	0,135	0,125	0,108	0,072	0,041	0,035	0,036
rW:	0,069	0,069	0,065	0,058	0,074	0,087	0,096
FP:	0,080	0,079	0,079	0,072	0,062	0,059	0,058
FPW:	0,057	0,056	0,055	0,051	0,048	0,049	0,049
FPC:	0,078	0,078	0,077	0,074	0,067	0,066	0,066
FPCW:	0,056	0,055	0,054	0,053	0,054	0,055	0,055
I. fajta hibabecslés átlagok ($\alpha = 0,10$)							
$(\sigma_1:\sigma_2)$	(4:1)	(3:1)	(2:1)	(1:1)	(1:2)	(1:3)	(1:4)
rt:	0,184++	0,170++	0,145++	0,100	0,074	0,071	0,072
rW:	0,120	0,115	0,109	0,103	0,114	0,129+	0,141++
FP:	0,129+	0,126+	0,122+	0,113	0,104	0,102	0,102
FPW:	0,102	0,101	0,098	0,093	0,091	0,092	0,093
FPC:	0,126+	0,125+	0,123+	0,119	0,114	0,113	0,112
FPCW:	0,099	0,099	0,099	0,099	0,100	0,102	0,102
I. fajta hibabecslés minimumok ($\alpha = 0,10$)							
$(\sigma_1:\sigma_2)$	(4:1)	(3:1)	(2:1)	(1:1)	(1:2)	(1:3)	(1:4)
rt:	0,169	0,151	0,121	0,081	0,068	0,067	0,068
rW:	0,114	0,108	0,103	0,098	0,102	0,114	0,127
FP:	0,124	0,122	0,117	0,107	0,101	0,100	0,100
FPW:	0,099	0,097	0,094	0,090	0,088	0,089	0,090
FPC:	0,122	0,121	0,119	0,115	0,111	0,110	0,110
FPCW:	0,097	0,096	0,096	0,096	0,097	0,099	0,099
I. fajta hibabecslés maximumok ($\alpha = 0,10$)							
$(\sigma_1:\sigma_2)$	(4:1)	(3:1)	(2:1)	(1:1)	(1:2)	(1:3)	(1:4)
rt:	0,203	0,196	0,176	0,129	0,085	0,077	0,076
rW:	0,132	0,126	0,118	0,110	0,136	0,153	0,163
FP:	0,137	0,133	0,130	0,120	0,110	0,106	0,105
FPW:	0,110	0,106	0,102	0,097	0,095	0,095	0,095
FPC:	0,132	0,129	0,127	0,123	0,118	0,115	0,115
FPCW:	0,105	0,102	0,101	0,101	0,103	0,104	0,105

Megjegyzés: A táblázatban a névleges szintet több mint 20, illetve 40%-os mértékkel meghaladó I. fajta hibabecslésátlagot +, illetve ++ jelöli.

9. táblázat. A vizsgált 6 próba ereje különböző mintaelemszámok mellett, 117 különböző eloszláspárra leátlagolva a $H_1: A_{12} = 0,64$ ellenhipotézis igaz volta esetén

Erőátlagok $\alpha = 0,05$, illetve $m = 6$ és $n = 12$ mellett							
$(\sigma_1:\sigma_2)$	(4:1)	(3:1)	(2:1)	(1:1)	(1:2)	(1:3)	(1:4)
rt:	(0,24)	(0,22)	(0,19)	0,16	0,14	0,13	0,14
rW:	(0,13)	(0,14)	(0,15)	0,17	(0,20)	(0,22)	(0,24)
FP:	(0,13)	(0,14)	(0,15)	0,15	0,14	0,14	0,14
FPW:	0,12	0,13	0,14	0,15	0,16	0,17	0,17
FPC:	(0,13)	(0,14)	(0,16)	0,16	0,16	0,15	0,15
FPCW:	0,12	(0,14)	(0,15)	0,17	0,18	0,18	0,18
Erőátlagok $\alpha = 0,10$, illetve $m = 6$ és $n = 12$ mellett							
$(\sigma_1:\sigma_2)$	(4:1)	(3:1)	(2:1)	(1:1)	(1:2)	(1:3)	(1:4)
rt:	(0,32)	(0,30)	(0,28)	0,25	0,23	0,23	0,23
rW:	0,23	0,22	0,23	0,25	0,29	(0,31)	(0,32)
FP:	(0,25)	(0,24)	0,24	0,24	0,24	0,24	0,24
FPW:	0,21	0,20	0,21	0,23	0,24	0,25	0,25
FPC:	(0,24)	(0,23)	0,24	0,26	0,27	0,26	0,26
FPCW:	0,16	0,18	0,20	0,25	0,27	0,28	0,28
Erőátlagok $\alpha = 0,05$, illetve $m = 12$ és $n = 24$ mellett							
$(\sigma_1:\sigma_2)$	(4:1)	(3:1)	(2:1)	(1:1)	(1:2)	(1:3)	(1:4)
rt:	(0,32)	(0,31)	(0,29)	0,27	0,25	0,25	0,25
rW:	(0,22)	(0,23)	0,24	0,29	0,35	(0,38)	(0,40)
FP:	(0,25)	(0,26)	(0,28)	(0,31)	0,33	0,33	0,33
FPW:	0,20	0,21	0,23	0,27	0,30	0,30	0,30
FPC:	(0,24)	(0,26)	(0,28)	(0,32)	(0,35)	(0,35)	(0,34)
FPCW:	0,20	0,21	0,23	0,28	0,32	0,32	0,32
Erőátlagok $\alpha = 0,10$, illetve $m = 12$ és $n = 24$ mellett							
$(\sigma_1:\sigma_2)$	(4:1)	(3:1)	(2:1)	(1:1)	(1:2)	(1:3)	(1:4)
rt:	(0,41)	(0,41)	(0,40)	0,38	0,37	0,37	0,37
rW:	(0,33)	0,33	0,35	0,40	0,46	(0,49)	(0,51)
FP:	(0,34)	(0,35)	(0,37)	0,41	0,44	0,44	0,43
FPW:	0,29	0,31	0,33	0,38	0,41	0,42	0,41
FPC:	(0,33)	(0,35)	(0,37)	0,42	0,45	0,46	0,45
FPCW:	0,29	0,30	0,33	0,39	0,43	0,44	0,43

Megjegyzés: A táblázatban zárójelbe tettünk minden olyan erőátlagot, amelyhez tartozó megfelelő I. fajta hibaátlag (vö. 3. és 4. táblázat) több mint 20%-kal meghaladja a névleges szintet.

10. táblázat. A vizsgált 6 próba ereje különböző mintaelemszámok mellett, 117 különböző eloszláspárra leátlagolva a $H_1: A_{12} = 0,71$ ellenhipotézis igaz volta esetén

Erőátlagok $\alpha = 0,05$, illetve $m = 6$ és $n = 12$ mellett							
$(\sigma_1:\sigma_2)$	(4:1)	(3:1)	(2:1)	(1:1)	(1:2)	(1:3)	(1:4)
rt:	(0,37)	(0,35)	(0,33)	0,30	0,29	0,28	0,28
rW:	(0,23)	(0,24)	(0,27)	0,32	(0,38)	(0,41)	(0,43)
FP:	(0,23)	(0,25)	(0,27)	0,30	0,30	0,29	0,29
FPW:	0,21	0,23	0,26	0,30	0,33	0,33	0,33
FPC:	(0,22)	(0,25)	(0,28)	0,31	0,32	0,31	0,31
FPCW:	0,22	(0,24)	(0,27)	0,33	0,35	0,35	0,35
Erőátlagok $\alpha = 0,10$, illetve $m = 6$ és $n = 12$ mellett							
$(\sigma_1:\sigma_2)$	(4:1)	(3:1)	(2:1)	(1:1)	(1:2)	(1:3)	(1:4)
rt:	(0,46)	(0,45)	(0,44)	0,43	0,42	0,42	0,42
rW:	0,36	0,36	0,37	0,43	0,50	(0,52)	(0,53)
FP:	(0,39)	(0,38)	0,39	0,42	0,44	0,44	0,44
FPW:	0,33	0,33	0,35	0,41	0,44	0,44	0,45
FPC:	(0,37)	(0,37)	0,39	0,44	0,46	0,46	0,46
FPCW:	0,26	0,29	0,34	0,43	0,47	0,48	0,48
Erőátlagok $\alpha = 0,05$, illetve $m = 12$ és $n = 24$ mellett							
$(\sigma_1:\sigma_2)$	(4:1)	(3:1)	(2:1)	(1:1)	(1:2)	(1:3)	(1:4)
rt:	(0,54)	(0,54)	(0,54)	0,55	0,54	0,54	0,54
rW:	(0,42)	(0,44)	0,48	0,57	0,66	(0,69)	(0,70)
FP:	(0,46)	(0,48)	(0,53)	(0,60)	0,64	0,63	0,63
FPW:	0,40	0,42	0,46	0,54	0,60	0,61	0,60
FPC:	(0,45)	(0,48)	(0,53)	(0,61)	(0,66)	(0,65)	(0,65)
FPCW:	0,39	0,42	0,46	0,56	0,62	0,63	0,62
Erőátlagok $\alpha = 0,10$, illetve $m = 12$ és $n = 24$ mellett							
$(\sigma_1:\sigma_2)$	(4:1)	(3:1)	(2:1)	(1:1)	(1:2)	(1:3)	(1:4)
rt:	(0,64)	(0,65)	(0,66)	0,67	0,68	0,68	0,68
rW:	(0,55)	0,57	0,60	0,69	0,76	(0,79)	(0,79)
FP:	(0,57)	(0,59)	(0,63)	0,70	0,74	0,74	0,73
FPW:	0,52	0,54	0,58	0,67	0,72	0,72	0,71
FPC:	(0,56)	(0,58)	(0,63)	0,71	0,75	0,75	0,75
FPCW:	0,51	0,54	0,58	0,68	0,73	0,74	0,73

Megjegyzés: A táblázatban zárójelbe tettünk minden olyan erőátlagot, amelyhez tartozó megfelelő I. fajta hibaátlag (vö. 3. és 4. táblázat) több mint 20%-kal meghaladja a névleges szintet.

Mindezen táblázatok alapján a következő megállapításokat tehetjük.

1. Kicsi és különböző nagyságú minták ($m = 6$, $n = 12$) mellett 5%-os szignifikanciaszinten csak egyetlen próba, FPW átlagos hibája marad minden szórásarány esetén a névleges szint (azaz 0,05) körüli 20%-os szélességű sávban (vö. 7. táblázat, legfelső blokk). A 117 eloszláspár I. fajta hibáit egyedileg szemlélve azt láthatjuk, hogy még a legszélsőségesebb esetben is csak 0,063 a valószínűsége a nullhipotézis jogtalan elutasításának, míg ugyanezen I. fajta hibamaximum az FPCW-, FP-, FPC- és rW-próba esetében rendre 0,068, 0,070, 0,072, illetve 0,093 (vö. 7. táblázat 3. blokkja). Minthogy ugyanezen elrendezésekben az FPW-próba ereje nem marad el számottevően a többi robusztus próba erejétől (vö. 9. és 10. táblázat legfelső blokkja), a legjobb próba ez esetben egyértelműen FPW.

A 0,06-nál nagyobb I. fajta hibát produkáló eloszláspárokat (ilyenből összesen pontosan 30 akadt) egyenként is megvizsgálva a következő érdekes eredmény adódott.

A második, vagyis a nagyobb elemszámú mintához tartozó 30 eloszlás közül 27 a maga ferdeségi szintjén a legalacsonyabb csúcossági értékű volt (vö. 1. táblázat), 3 eloszlás csúcossága pedig a középső övezetbe esett. Ugyanakkor a ferdeség tekintetében 8-12-10 volt a három ferdeségi szint közötti megoszlás, vagyis az eloszlás szimmetrikus-aszimmetrikus volta nem hatott az I. fajta hiba szintjére. Ez az eredmény erősíti azt a nézetet, hogy a statisztikai próbák érvényességét nemcsak az eloszlás ferdesége, hanem esetenként a csúcossága is befolyásolja (vö. VARGHA, 1996). Ez egyébként összhangban van azokkal a vizsgálatokkal is, amelyek magas csúcossági szintű eloszláspárok esetén a kétmintás t - és a Welch-próba alacsony erejét, egyben túlzott konzervativizmusát (az I. fajta hiba túlzottan alacsony szintjét) mutatták ki (vö. WILCOX, 1996, 148–154), ami ugyanannak a jelenségnek az ellentétes oldala (alacsony csúcosság – magas I. fajta hiba, magas csúcosság – alacsony I. fajta hiba).

2. Ugyanezen elrendezésekben $\alpha = 10\%$ esetén FPW mellett FPCW I. fajta hibabecslésátlaga marad még minden szórásarány esetén a 20%-os eltéréssávon belül. A legnagyobb eltérést a 4:1 szórásarány idézi elő. Ez esetben érdekes módon az FPW-próba I. fajta hibája felfelé leng ki (minimum = 0,104, maximum = 0,132), az FPCW-próbáé pedig lefelé (minimum = 0,067, maximum = 0,090; vö. 7. táblázat alsó két blokkjának első adatoszlópa). Némileg hasonló a helyzet a fordított, 1:4 szórásarány mellett is, csak ellenkező irányban. Itt ugyanis az FPW-próba I. fajta hibája lefelé leng ki (minimum = 0,083, maximum = 0,092), az FPCW-próbáé viszont felfelé (minimum = 0,100, maximum = 0,110; vö. 7. táblázat alsó két blokkjának utolsó adatoszlópa). Az erőértékeket is figyelembe véve (lásd 9. és 10. táblázat 2. blokkja) azt állapíthatjuk meg, hogy azonos irányú minta-elemszám- és szórásarány, illetve $\sigma_1 = \sigma_2$ esetén FPCW, ellenkező esetben pedig FPW tűnik a legjobb robusztus próbának.

Megint csak megnézve egyenként a legmagasabb I. fajta hibákat produkáló eloszláspárokat, az az eredmény adódott, hogy a 12 darab 0,120 fölötti I. fajta hibát eredményező eloszláspár esetében az első, vagyis a kisebb elemszámú mintához tartozó eloszlás 9 esetben alacsony csúcosságú szimmetrikus, 3 esetben pedig

alacsony csúcosságú, közepesen ferde volt, tehát ebben az esetben is az egyik minta alacsony csúcosságú eloszlásai okozták az I. fajta hiba megnövekedését.

3. Közepes és eltérő nagyságú minták ($m = 12$, $n = 24$) esetén 5%-os szignifikanciaszinten FPW és FPCW imponálón szép I. fajta hibákat produkál, miközben a többi próba I. fajta hibaszintje számos esetben megengedhetetlenül magasra kúszik (vö. 8. táblázat 1. blokkja). Tekintve, hogy FPCW ereje esetenként (azonos irányú elemszámarány és szórásarány mellett) 1-2 százalékponttal nagyobb, mint FPW ereje (vö. 9. és 10. táblázat 3. blokkja), ez esetben FPCW enyhén jobbnak tűnik, mint FPW, bár ez utóbbi is igen jó teljesítményt mutat.

4. Ugyanezen elrendezésekben $\alpha = 10\%$ esetén ugyanazt állapíthatjuk meg, mint 5%-os szignifikanciaszinten: FPW és FPCW egyaránt megfelelő próbának tűnik, FPCW enyhe fölényével (vö. 8. táblázat alsó fele, illetve 9. és 10. táblázat legalsó blokkja).

A fenti konklúziók eredőjeként azt állapíthatjuk meg, hogy különböző elemszámú minták esetén, ha a minták kicsik, FPW tűnik a sztochasztikus egyenlőség legjobb próbájának, mert az összes többi bizonyos elrendezésekben a névlegest lényegesen meghaladó I. fajta hibát produkál. Közepes nagyságú minták esetén viszont FPCW felzárkózik FPW mellé, sőt, esetenkénti erőfölénye miatt FPW-nél csekély mértékben jobbnak is tűnik.

Az FP-próba a várakozással ellentétben nem bizonyult eléggé robusztusnak a szórásheterogenitással szemben. Például $m = 6$ és $n = 12$, illetve $\alpha = 5\%$ mellett az I. fajta hibaszintje 4:1, 3:1 és 2:1 szórásarány esetén egyaránt 0,065 körül mozog, az 1:3 és 1:4 arány esetén viszont 0,037-re esik le (vö. 7. táblázat legfelső blokkja). Hasonló irányú viselkedés figyelhető meg $\alpha = 10\%$ mellett is (vö. 7. táblázat 4. blokkja). Közepes minták esetén szintén az elemszámarányokkal fordított irányú szórásarányok vezetnek túlságosan megemelkedett I. fajta hibaszinthez (vö. 8. táblázat 1. és 4. blokkja).

Az FPC-próba különböző nagyságú minták esetén végig FP-hez nagyon hasonlóan viselkedik.

Az rW-próba legfeljebb kétszeres elméleti szóráseltérés (2:1, 1:1, 1:2 szórásarány) esetén elfogadható érvényességű. Nagyobb mértékű szórásheterogenitás esetén sajnos túlzottan magas I. fajta hibák jellemzik.

Az rt-próba – a kétmintás t-próbához hasonlóan – különböző mintaelemszámok esetén csak azonos elméleti szórások mellett tekinthető érvényes eljárásnak. A szórások közti akár csak 1:2, illetve 2:1 arányú eltérés esetén is I. fajta hibája elfogadhatatlan mértékben különbözhet a névleges szinttől.

4. MEGBESZÉLÉS

A jelen tanulmányban a sztochasztikus egyenlőség nullhipotézisét tesztelő hat nemparaméteres statisztikai próbát hasonlítottunk össze számítógépes szimulációval az érvényesség és a hatékonyság szempontjából. Az elméleti átlagok, mediánok, illetve más középértékek egyenlőségével szemben a sztochasztikus egyenlőség nem a két populáció egy-egy jellegzetes paraméterének ugyanakkorosságát állítja, ha-

nem egészes összehasonlításban jelez egyfajta egyenértékűséget a két populáció között. Két populációt akkor mondunk sztochasztikusan egyenlőnek valamely X változó tekintetében, ha véletlenszerűen kiválasztva egy-egy X -értéket a két populációból, az elsőből kiválasztott érték ugyanakkora eséllyel lesz nagyobb a másodikból kiválasztottnál, mint kisebb.

A sztochasztikus egyenlőség vizsgálatára korábban több robusztus, az eloszlások alakjára, illetve a szórás heterogenitásra kevésbé érzékeny próbát is javasoltak (FLIGNER, POLICELLO, 1981; CLIFF, 1993; VARGHA, 1999a). Ezen eljárások alkalmaságát azonban mindeddig nem támasztották alá olyan alapos szimulációs vizsgálatok, amelyek az eloszlások, elméleti szórások és elemszámok különböző arányaira, illetve kombinációira hitelesítették, illetve összevetették volna a próbákat. A jelen tanulmány ezt a feladatot vállalta fel.

A szimuláció során széles tartományban variáltuk az eloszlások ferdeségét és csúcosságát, összevetve szimmetrikus, továbbá azonos, illetve különböző irányban ferde eloszlásokat. A szórás heterogenitást négy szinten ($\sigma_1:\sigma_2 = 1:1, 1:2, 1:3, 1:4$), a minták egyenlőségét (egyenlő, illetve különböző) és nagyságát (kicsi, illetve közepes) két-két szinten vizsgáltuk. A szimulációba a korábban már mások által is javasolt négy próba (rt, rW, FP és FPC) mellett elméleti megfontolások alapján két új próbát (FPW és FPCW) is bevontunk. FP és FPC, illetve FPW és FPCW tesztstatisztikája egymással megegyezik, csak tesztkiértékelésük módja tér el egymástól. Előbbiek (FP és FPC) kis minták esetén egy speciális táblázat (vö. FLIGNER, POLICELLO, 1981, 1. táblázat vagy WILCOX, 1996, 425), nagy minták esetén pedig a standard normális eloszlás táblázata alapján értékelhetők ki. Utóbbiak (FPW és FPCW) a t-eloszlás táblázata alapján, egy olyan speciális szabadságfok felhasználásával értékelhetők ki, amelynek formulája azonos a Welch-próbában alkalmazott szabadságfok formulájával (lásd a Melléklet (7) formuláját).

A szimulációs vizsgálatok arra az érdekes eredményre vezettek, hogy az újonnan javasolt két próba, FPW és FPCW az érvényesség tekintetében sokkal megbízhatóbb eljárásnak bizonyult, mint a többiek, miközben az erő tekintetében nem tapasztaltunk számottevő különbséget közöttük. Különösen FPW jeleskedett azzal, hogy I. fajta hibája sosem tért el számottevően a névleges szinttől. Például 5%-os szignifikanciaszint mellett az összesen megvizsgált $4 \times 7 \times 117 = 3276$ elrendezésben FPW I. fajta hibája kis minták esetén a (0,043–0,063), közepes minták esetén pedig a (0,043–0,057) szűk sávban ingadozott. Összehasonlításképpen rW, FP, FPC és FPCW I. fajta hibája kis minták esetén rendre a (0,049–0,093), (0,035–0,070), (0,040–0,072), (0,049–0,068), közepes minták esetén pedig rendre a (0,050–0,096), (0,054–0,080), (0,061–0,078), (0,049–0,058) sávban ingadozott.

FPW jósága abból is megítélhető, hogy az elméleti átlagok összehasonlítására két független minta esetén használt népszerű Welch-féle robusztus próba I. fajta hibája 5%-os szinten esetenként (pl. lognormális eloszlású függő változókat választva) még 100-as összelemszámnál is meghaladhatja a névleges szint 200%-át, azaz 0,1-et (például $m = 65, n = 35, \sigma_1 = 1, \sigma_2 = 3$ esetén), illetve még 500-as összelemszámnál is meghaladhatja a 0,064-et (például $m = 325, n = 175, \sigma_1 = 1, \sigma_2 = 3$ esetén), ahogy ezt egy nemrég megjelent tanulmány bemutatta (ALGINA, OSHIMA, LIN, 1994, 2. táblázat).

Közepes nagyságú minták esetén FPCW FPW-vel egyenértékű jó eljárásnak tűnik. Vele kapcsolatban további szimulációs vizsgálatokkal kellene kideríteni, hogy pontosan mi az a mintanagyság, ami mellett I. fajta hibája nem emelkedik 20%-osnál nagyobb mértékben a névleges szint fölé.

Vizsgálatunkban különböző mintaelemszámok esetén az 1:2 elemszamarányt alkalmaztuk. Mivel az empirikus pszichológiai vizsgálatokban időnként ennél szélsőségesebb elemszamarányok is előfordulnak, jó lenne a szimulációt ilyen esetekre is elvégezni.

A szimulációban alkalmazott lambdaeloszláscsalád a folytonos eloszlások valóban széles spektrumát fedi le (vö. VARGHA, 1996, 1. ábra). Hiányoznak azonban közülük a jellegzetes kétsúcú, bimodális eloszlások, amelyeket például két normális eloszlás alkalmas „keverésével” lehet előállítani. A szimulációs vizsgálatokat jó lenne ilyen irányban is kiterjeszteni, s hasonló javasolhatunk a diszkrét eloszlásokkal kapcsolatban is, amelyek a pszichológiában különösen gyakoriak (vö. MICCERI, 1989).

Végső konklúzióként megállapíthatjuk, hogy a sztochasztikus egyenlőség nullhipotézisének vizsgálatára sikerült egy régebben javasolt nemparaméteres statisztikai próbát (vö. FLIGNER, POLICELLO, 1981) úgy módosítani (FPW-próba), hogy az néhány kis és közepes elemszám-kombináció esetén az eloszlások és szórásarányok igen széles tartományában imponálóan kedvező I. fajta hibát produkáljon, érezhetően jobbat, mint az eddig ugyanerre a célra javasolt más eljárások. Hasonló állapítható meg közepes minták esetén az FPCW-próbáról is, mely ezekben az elrendezésekben FPW-vel egyenértékű jó eljárásnak bizonyult.

Végül megjegyezzük, hogy FPW és FPCW segítségével a valószínűségi fölény A mutatójára az eddig ismertnél pontosabb intervallumbecslés is készíthető (a módszer technikai részleteivel kapcsolatban lásd VARGHA, 1999a, 4. pont) és hogy az A mutató becslése, illetve a sztochasztikus egyenlőség tesztelése könnyűszerrel elvégezhető a szerző által kidolgozott MiniStat statisztikai programcsomag legújabb változata segítségével (VARGHA, 1999b).

MELLÉKLET

A szimulációs elemzésbe bevont statisztikai próbák leírása

Tételezzük fel, hogy egy legalább ordinális skálájú változó segítségével szeretnénk összehasonlítani két populációt, Pop1-et és Pop2-t. Jelöljük ezt a változót Pop1-ben X, Pop2-ben pedig Y betűvel. A szimulációs elemzésbe bevont próbák mindegyikével Pop1 és Pop2 sztochasztikus egyenlőségét ellenőrizzük, vagyis a

$$H_0: P(X > Y) = P(X < Y) \quad (1)$$

nullhipotézist, mely ekvivalens a valószínűségi fölény A mutatójára vonatkozó

$$H_0: A_{12} = 0,5 \quad (2)$$

hipotézissel (vö. a cikkünk bevezetőjében mondottakkal).

Legyen X_1, X_2, \dots, X_m egy Pop1-ből véletlenszerűen kiválasztott m elemű X -minta, Y_1, Y_2, \dots, Y_n pedig egy Pop2-ből véletlenszerűen és az előzőtől függetlenül kiválasztott n elemű Y -minta. Ekkor a sztochasztikus egyenlőség (1) nullhipotézisét a szimulációba bevont hat statisztikai próbával a következőképpen lehet megvizsgálni.

1. Rang t -próba (rt)

Rangsoroljuk nagyság szerint az X_1, X_2, \dots, X_m és az Y_1, Y_2, \dots, Y_n adatokat közös mintában úgy, ahogy az a Mann-Whitney-próbában szokás (vö. HAJTMAN, 1968, 348). A kapott rangszámokat a két mintára vonatkozóan jelölje rendre r_1, r_2, \dots, r_n , illetve q_1, q_2, \dots, q_n . Ezután e két rangszámminta felhasználásával végezzük el a kétmintás t -próbát (vö. HAJTMAN, 1968, 159). Ha ez a rangszámokon végrehajtott kétmintás t -próba szignifikáns, akkor utasítsuk el a két populáció sztochasztikus egyenlőségét állító (1) nullhipotézist.

2. Rang Welch-próba (rW)

Ez a próba csak annyiban különbözik az előzőleg bemutatott rt -próbától, hogy a rangsorolás eredményeként kapott r_1, r_2, \dots, r_n és q_1, q_2, \dots, q_n rangszámmintán nem a kétmintás t -, hanem a Welch-féle d -próbát (vö. HAJTMAN, 1968, 166) hajtjuk végre.

3. Fligner–Policello-próba (FP)

Minden X - és Y -mintabeli adathoz számítsunk ki egy ún. *helyértéket* (*placement score*), melyet jelöljön V_i ($i = 1, \dots, m$) az X -minta és W_j ($j = 1, \dots, n$) az Y -minta esetében, a következő módon. Legyen V_i mindazon Y -adatok száma, amelyek kisebbek, mint X_i ($i = 1, \dots, m$), és hasonlóképpen legyen W_j mindazon X -adatok száma, amelyek kisebbek, mint Y_j ($j = 1, \dots, n$). Ezen jelölések mellett az FP -próba statisztikája a

$$Z_{FP} = d/s_d, \quad (3)$$

mennyiség, ahol a d számláló a $\delta = P(X > Y) - P(X < Y)$ sztochasztikus különbség alábbi pontbecslése:

$$d = (\sum_i V_i - \sum_j W_j)/(mn), \quad (4)$$

az s_d nevező pedig ezen d mennyiség elméleti szórásának egy pontbecslése, mely a $\bar{v} = \sum_i V_i/m$, $\bar{w} = \sum_j W_j/n$, $Q_V = \sum_i (V_i - \bar{v})^2$, $Q_W = \sum_j (W_j - \bar{w})^2$ jelölésekkel a következőképpen írható fel:

$$s_d = 2(Q_V + Q_W + \bar{v}\bar{w})^{1/2}/(mn). \quad (5)$$

Az FP-próba azon alapul, hogy Z_{FP} tesztstatisztikája aszimptotikusan (a minta-elemszámok növelésével) standard normális eloszlást követ, tehát Z_{FP} szignifikanciáját a standard normális eloszlás táblázata alapján lehet kiértékelni. Fligner és Policello azonban 3 és 12 közti elemszámokra "egzakt" kritikus értékeket is közölt (lásd FLIGNER, POLICELLO, 1981, 1. táblázat, illetve WILCOX, 1996, B Melléklet, 14. táblázat), így a szimulációban kis minták ($m + n = 18$) esetén ezeket alkalmaztuk.

4. A Fligner–Policello-próba Cliff-féle módosítása (FPC)

CLIFF (1993) alternatív képletet javasolt a Z_{FP} mennyiségre. Véleménye szerint a fenti (4) formulával megadott d mennyiség elméleti szórását az (5) képlettel megadott s_d helyett jobb lenne az alábbi formula által meghatározott S_d mennyiséggel becsülni:

$$(S_d)^2 = \frac{n^2 \sum_i (d_i - d)^2 + m^2 \sum_j (d_j - d)^2 + \sum_i \sum_j (d_{ij} - d)^2}{mn(m-1)(n-1)} \quad (6)$$

(lásd CLIFF, 1993, (9) formula). Itt $d_{ij} = \text{sign}(X_i - Y_j)$, $d_i = \sum_j d_{ij}/n$, $d_j = \sum_i d_{ij}/m$ és d , az összes d_{ij} érték átlaga ugyanaz, mint a (4) formula által meghatározott mennyiség (a $\text{sign}(c)$ előjelváltozó értéke bármely c szám esetén -1 , 0 vagy 1 attól függően, hogy c negatív, nulla vagy pozitív). FPC szignifikanciájának megállapítása ugyanúgy történik, mint az FP-próbáé, vagyis kis minták ($m, n \leq 12$) esetén az FP-próba kritikus értékeinek táblázata alapján, nagyobb minták esetén pedig a standard normális eloszlás táblázata segítségével.

5. FP-próba Welch-féle szabadságfokkal kiértékelve (FPW)

Ennek a próbának ugyanaz a próbastatisztikája, mint FP-é (Z_{FP}), csak ennek szignifikanciáját nem az FP-próba esetében említett táblázatok, hanem a t-eloszlás táblázata segítségével értékeljük ki. A t-táblázat használatához szükséges megfelelő f szabadságfokot a Welch-próbában alkalmazott módon számítjuk ki oly módon, hogy nem az eredeti adatok, hanem az FP-próba esetében definiált helyértékek X - és Y -mintára vonatkozó varianciáját vesszük alapul. Képlettel kifejezve:

$$f = \frac{(a + b)^2}{a^2/(m-1) + b^2/(n-1)}, \quad (7)$$

ahol

$$a = \frac{Q_v}{(m-1)m} \quad \text{és} \quad b = \frac{Q_w}{(n-1)n}$$

(itt m és n a két minta elemszáma, Q_v és Q_w pedig az FP-próba esetében fentebb már definiált mennyiségek). A (7) formula természetesen rendszerint nem egész értéket szolgáltat, így az f szabadságfok végső értéke a (7) formulával kapott érték egész számra való kerekítésével adódik.

6. FPC-próba Welch-féle szabadságfokkal kiértékelve (FPCW)

Ez a próba csak abban különbözik az előző pontban bemutatott FPW-próbától, hogy tesztstatisztikáját nem az FP, hanem az FPC-próbában leírt módon képezzük, vagyis az FPC-próba statisztikáját a (7) formulával megadott szabadságfokú t -eloszlás segítségével értékeljük ki.

IRODALOM

- ALGINA, J., OSHIMA, T. C., LIN, W. Y. (1994) Type I error rates for Welch's test and James's second-order test under nonnormality and inequality of variance when there are two groups. *Journal of Educational and Behavioral Statistics*, 19, 275–291.
- CLIFF, N. (1993) Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin*, 114, 494–509.
- CONOVER, W. J., IMAN, R. L. (1981) Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician*, 35, 124–129.
- FLIGNER, M. A., POLICELLO II, G. E. (1981) Robust rank procedures for the Behrens-Fisher problem. *Journal of the American Statistical Association*, 76, 323–327.
- HAJTMAN B. (1968) *Bevezetés a matematikai statisztikába pszichológusok számára*. Akadémiai Kiadó, Budapest
- MCKEAN, J. W., VIDMAR, T. J. (1994) A comparison of two rank-based methods for the analysis of linear models. *The American Statistician*, 48, 220–229.
- MICCERI, T. (1989) The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, 105, 156–166.
- ONGHENA, P. (1993) A theoretical and empirical comparison of mainframe, microcomputer, and pocket calculator pseudorandom number generators. *Behavior Research Methods, Instruments, & Computers*, 25, 384–395.
- RAMBERG, J. S., TADIKAMALLA, P. R., DUDEWICZ, E. J., MYKYTKA, E. F. (1979) A probability distribution and its uses in fitting data. *Technometrics*, 21, 201–209.
- SCHEFFÉ, H. (1959) *The analysis of variance*. Wiley, New York
- VARGHA A. (1996) Az egymintás t -próba érvényessége és javíthatósága. *Magyar Pszichológiai Szemle*, LII (36), 4–6, 317–345.

- VARGHA A. (1999a) Két csoport összehasonlítása nemparaméteres statisztikai eljárások segítségével. *Magyar Pszichológiai Szemle*, 54, 567–589.
- VARGHA A. (1999b) *MiniStat 3.1 verzió. Felhasználói füzet*. Pólya Kiadó, Budapest
- VARGHA A. (2000) *Matematikai statisztika pszichológiai, nyelvészeti és biológiai alkalmazásokkal*. Pólya Kiadó, Budapest
- VARGHA, A., DELANEY, H. D. (1998) The Kruskal-Wallis test and stochastic homogeneity. *Journal of Educational and Behavioral Statistics*, 23, 170–192.
- VINCZE I. (1968) *Matematikai statisztika ipari alkalmazásokkal*. Műszaki Könyvkiadó, Budapest
- WILCOX, R. R. (1996) *Statistics for the social sciences*. Academic Press, San Diego–New York
- ZIMMERMAN, D. W., ZUMBO, B. D. (1992) Parametric alternatives to the Student t test under violation of normality and homogeneity of variance. *Perceptual and Motor Skills*, 74, 835–844.
- ZIMMERMAN, D. W., ZUMBO, B. D. (1993a) Rank transformations and power of the Student t test and Welch t-test for non-normal populations with unequal variances. *Canadian Journal of Experimental Psychology*, 47, 523–539.
- ZIMMERMAN, D. W., ZUMBO, B. D. (1993b) The relative power of parametric and non-parametric statistical methods. In Keren, G., Lewis, C. (eds) *A handbook for data analysis in the behavioral sciences: Methodological issues*. 481–517. Erlbaum, Hillsdale
- ZUMBO, B. D., COULOMBE, D. (1997) Investigation of the robust rank-order test for non-normal populations with unequal variances: The case of reaction time. *Canadian Journal of Experimental Psychology*, 51, 139–149.

THE COMPARISON OF SEVERAL TESTS OF STOCHASTIC EQUALITY

VARGHA, ANDRÁS

In the current paper six statistical tests of stochastic equality are to be compared by a Monte Carlo simulation with respect to Type I error and power. Two populations are said to be stochastically equal with respect to a variable X, if for any two independently and randomly drawn observations X_1 and X_2 from the two populations $P(X_1 > X_2) = P(X_1 < X_2)$.

In the simulation the skewness and kurtosis levels as well as the extent of variance heterogeneity of the two parent distributions were varied across a wide range. The sample sizes applied were either small or moderate, and equal or unequal. The involved tests of stochastic equality were as follows: rank t test, rank Welch test, Fligner–Policello test, Cliff's modified Fligner–Policello test as well as two modifications of the last two tests, denoted FPW and FPCW, that utilized adjusted degrees of freedom.

An interesting result obtained is that the two newly introduced test variants, FPW and FPCW, proved to be substantially more accurate with regard to their Type I error rates than the others, whereas they kept a similar power level. Specifically, the estimated Type I error of FPW at .05 nominal level always fell in the range of .043–.063 even if the variance ratio of the two distributions was as large as 1:16. The same ranges were .049–.068 for FPCW, but .029–.160 for the rank t test, .049–.096 for the rank Welch test, .035–.075 for the Fligner–Policello test, and .040–.078 for Cliff's test.

Key words: *group comparison, stochastic equality, stochastic difference, measure of stochastic superiority, rank Welch test, Fligner–Policello test, Cliff's modified Fligner–Policello test, FPW test, FPCW test*