

**FEJLESZTŐPROGRAMOK ÖNKONTROLLÓS HATÁSVIZSGÁLATÁNAK  
MATEMATIKAI STATISZTIKAI HÁTTERE**

**Szerzők:**

Mező Ferenc  
Debreceni Egyetem

Máth János  
Debreceni Egyetem

Abari Kálmán  
Debreceni Egyetem

Mező Katalin  
Debreceni Egyetem

Első szerző e-mail címe:  
ferenc.mezo1@gmail.com

**Lektorok:**

Demetrovics János  
Eötvös Lóránd Tudományegyetem

Nagy Dénes  
Nemzetközi Szimmetria Társaság

Varga Imre  
Szegedi Tudományegyetem

Koncz István  
Professzorok az Európai Magyarországiért

Mező F., Máth J., Abari K. és Mező K. (2015): Fejlesztőprogramok önkontrollos hatásvizsgálatának matematikai statisztikai háttere. *Különleges Bánásmód*, I. évf. 2015/4. szám, 65-75. DOI 10.18458/KB.2015.4.65

**Absztrakt**

*E tanulmány a különleges bánásmódot igénylő tanulók számára készült fejlesztőprogramok (elővizsgálat-fejlesztés-utóvizsgálat elrendezésű) önkontrollos hatásvizsgálatának matematikai statisztikai elemzéséhez nyújt módszertani útmutatót. A szükséges alapfogalmak össze-foglalása után bemutatjuk, hogy egy, az adatelemzéshez használható, ingyenesen letölthető szoftver (az R-nyelv) számára milyen utasításokat adhatunk, s a szoftver által elvégzett statisztikai műveletek eredményeit miként értelmezhetjük.*

**Kulcsszavak:** fejlesztőprogram, hatásvizsgálat, statisztika

**Diszciplínák:** matematika, pszichológia, gyógypedagógia, pedagógia

**Abstract**

*MATHEMATICAL STATISTICAL BACKGROUND OF SELF-CONTROLLED EFFECTIVENESS STUDY OF DEVELOPMENT PROGRAMMES*

*This paper offers a methodological guidance to mathematical statistical analysis of self controlled (pre-test - development - post-test type) effectiveness studies of development programmes written for children who need special treatment. After summarising the basic*

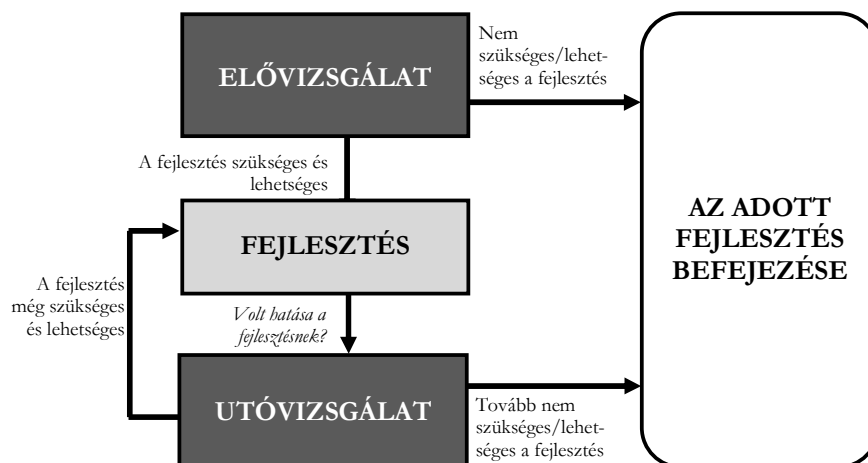
*terms we show what kind of commands can be given to a downloadable free software (it is the R language), and how we can interpret the results of statistical calculations done by the software.*

**Keywords:** development programmes, effectiveness study, statistics, R language

**Disciplines:** mathematics, psychology, special education, pedagogy

E tanulmányban a „fejlesztőprogram” kifejezésen olyan (pedagógiai, gyógypedagógiai, fejlesztőpedagógiai, pszichológiai orientációjú) beavatkozást értünk, mely tervszerűen (ideális esetben egy írásban rögzített fejlesztési terv alapján) igyekszik egy kedvezőtlenebbnek tartott értékről kedvezőbb irányba változtatni a program alanyainak tekinthető személyek valamely jól körülírt jellemzőjét (pl. figyelmét, önismeretét, tanulási stílusát stb.). E programokban/tervekben a fejlesztési cél meghatározása legalább kétféle lehet: a) konkrét cél lehet egy előre meghatározott kritériumszint (pl. ponthatár) elérése (bővebben lásd: Mező és tsai, 2015); b) általánosabb célként határozható meg, hogy a fejlesztés előtt és után végzett vizsgálat eredményei között különbség mutakozzon az utóvizsgálat javára. Jelen tanulmányban az utóbbi, általánosabb jellegű célmeghatározás ellenőrzésére szolgáló önkontrollos hatásvizsgálatra fókuszálunk a továbbiakban. A fejlesztőprogramok hatásának „önkontrollos” jellege arra utal, hogy a program alanyainak tekinthető személyek önmagukhoz képest mért fejlődésére fókuszálunk egy elővizsgálat-fejlesztőprogram-utóvizsgálat jellegű folyamatban (1. ábra).

1. ábra: az önkontrollos hatásvizsgálat sémája (forrás: a Szerzők)



Egy fejlesztőtevékenység önkontrollos hatásvizsgálata során az alábbiakhoz hasonló kérdésekre kell választ adnunk:

- Az „IPOO – minimum program” tanulás módszertani tréning előtt és után mért tanulás módszertani teljesítményben (a Jupiterbolha-próba összpontszámában) jelentős különbség tapasztalható-e? Különösen: kedvező (teljesítményt növelő) hatása volt-e a tréningnek, vagy sem?
- Igaz-e, hogy a csoport félév végi matematika felmérő dolgozatra kapott érdemjegyei lényeges változást mutatnak a félév eleji (hasonló tartalmú és nehézségű) dolgozathoz képest?
- Egy fejlesztő program elején és végén a Kozéki-féle „Iskolai orientáció” kérdőív révén felmérjük, hogy a diákok tanulási orientációja közül melyik a legdominánsabb kategória: a mélyreható tanulás, a reprodukatív tanulás vagy a szervezett tanulás? Vizsgálatunkban arra voltunk kíváncsiak, hogy a tehetség-gondozó programban való részvétel megváltoztatta-e a diákok tanulási orientációjának jellegét.

E kérdések közös vonása tehát, hogy egy csoport két vizsgálati alkalommal nyert eredményét hasonlítjuk össze (önkontrollos hatásvizsgálatot végzünk). Csoportos fejlesztések esetében az átlagos pontszámok, százalékos értékek számításán túl matematikai statisztikai különbségvizsgálatokkal támaszthatjuk alá vagy cáfolhatjuk meg a fejlesztőprogram hatásával kapcsolatos hipotéziseinket. E hipotézisek vizsgálata rámutathat arra, hogy a csoportban elért fejlesztőhatás szignifikáns (nem véletlenszerű) mértékű-e, vagy csupán eseti jelenségnek tekinthető-e.

Az alábbiakban olvasható útmutató a hipotézisvizsgálat - R statisztikai szoftverrel végezhető - adatelemzéséhez kíván módszertani segítséget nyújtani, a bemutatott statisztikai eljárások elméleti háttérének tárgyalása azonban nem célunk. A felmerülő elméleti kérdések tárgyalására a következő, magyarul is elérhető irodalmakat javasoljuk: Vargha (2000), Reiczigel és mtsai (2007), Máth (2004), Falus és Ollé (s.a.), Mező, Máth és Abari (2008), Máth és tsai (2015), továbbá az R adatelemző szoftverrel kapcsolatos praktikus ismereteket Abari és tsai (2015) írása foglalja össze.

Jelen tanulmányban az R parancsokkal kapcsolatban az Abari (2008) által megadott konvenciókat követjük. Az alábbi példákban szereplő R parancsok ezért alapesetben egy  $d$  nevű adattáblát használnak, amelyben minden megfigyelt személy adata egyetlen sorban szerepel, még akkor is, ha több különböző mérés tartozik hozzá:

	X1	X2
1	102	102
2	98	105
stb...	stb...	stb...
10	100	103

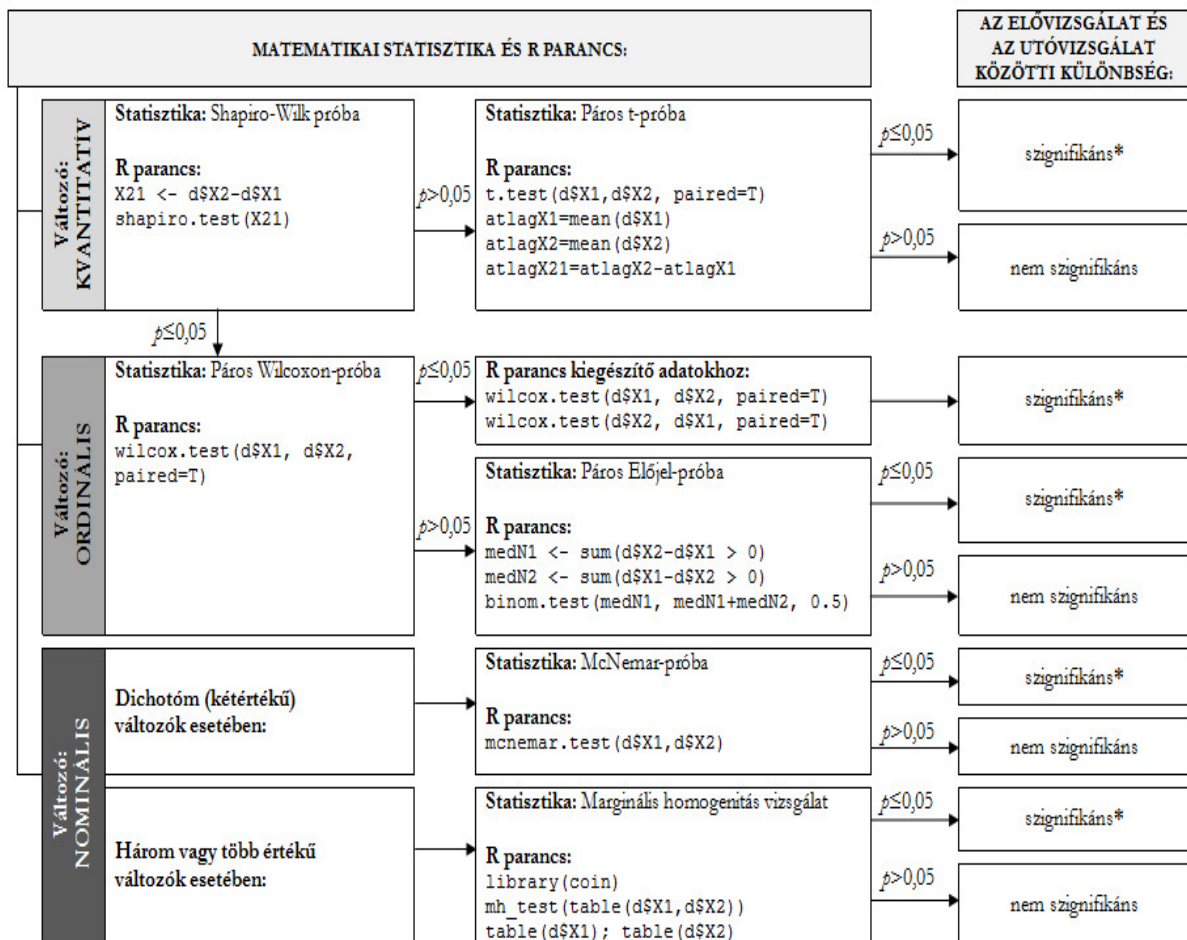
Az adattábla oszlopai a vizsgált változók (pl. X1, X2 vagy pl. az „IQ-elővizsgálat”, „IQ-utóvizsgálat” változók) adatait tartalmazzák. Az azonos személyek különböző mérési időpontok közötti összetartozását  $X1$ ,  $X2$  nevekkel jelöljük, így ezek a változók az R parancsokban  $d\$X1$ , illetve  $d\$X2$  módon jelennek meg (pl.  $X1$  = előtesztbeli teljesítmény;  $X2$  = utótesztbeli teljesítmény; a  $d\$X1$  jelentése pedig:  $d$  adatbázis  $X1$  változójára hivatkozunk az R parancsban, stb.).

### Statisztikaválasztás és végrehajtás az R-nyelvben

A matematikai statisztika elemzés kiindulópontja egy olyan adatbázis létrehozása, amelyben egy oszlopban (legyen a neve:  $X1$  oszlop) felsoroljuk a fejlesztés előtti eredményeket, s egy másik oszlopban (legyen most a neve ennek:  $X2$ ) a fejlesztést követő eredményeket soroljuk fel. Lényeges, hogy egy tanuló adatai egy sorba kerüljenek! Az adatbázis létrehozását követően ki kell választanunk az adatainknak megfelelő statisztikai eljárást (szokás statisztikai próbának is nevezni ezeket).

A megfelelő statisztikai próba kiválasztásának menete az önkontrollos hatásvizsgálat során (vö.: 2. ábra):

2. ábra: Statisztikaválasztás önkontrollos hatásvizsgálat esetében ( $d\$X1$  = elővizsgálat és  $d\$X2$  = utóvizsgálat adatait tartalmazó változók a „d” adatbázisban). Forrás: Mező, Máth és Abari (2008) alapján Mező F.



\* Ha a változó magasabb értéke jelenti a kedvezőbb eredményt (pl. jobb képességet), akkor fejlesztő hatásról  $d\$X1 < d\$X2$ ; visszafejlődésről  $d\$X1 > d\$X2$  viszony esetében beszélhetünk. Ha a változó alacsonyabb értéke jelenti a kedvezőbb eredményt (pl. kisebb mértékű szorongást), akkor  $d\$X1 > d\$X2$  eredmény jelenti a fejlődést, s a  $d\$X1 < d\$X2$  eredmény informál kedvezőtlen változásról.

1. meghatározzuk, hogy nominális, ordinális vagy kvantitatív változókkal kell-e számolnunk a továbbiakban. Értékeik jellegzetességei alapján ugyanis három változótípust különböztethetünk meg:
  - nominális változót, melynek egymástól megkülönböztethető, de nem rangsorolható értékei vannak;
  - ordinális változót, melynek értékei nemcsak megkülönböztethetők, de a kisebb-nagyobb dimenzió mentén rangsorolhatók, ugyanakkor két-két értéke közötti különbség nem egységes (pl. az iskolai osztályzatok esetében az elégtelen és az elégséges érdemjegy közötti távolság általában jóval nagyobb, mint a jó és jeles közötti távolság, amikor százalékos értékekben fejezik ki a pedagógusok az érdemjegyek ponthatárait);
  - kvantitatív változót (értékei nemcsak megkülönböztethetők és rangsorolhatók, de két értékük különbsége is értelmezhető);
2. kvantitatív változók esetében szoftver segítségével döntjük el, hogy normális eloszlásúnak tekinthető (harang alakú görbét mutató) változóról van-e szó;
3. az előző két lépés alapján kerülhet megválasztásra, végrehajtásra az adekvát statisztikai próba;
4. olykor további elemzésekre is szükség lehet.

A 2. ábra mutatja be, hogy milyen döntések meghozatala után választhatjuk ki az adekvát matematikai statisztikai eljárást, amennyiben egy vizsgálati csoport két alkalommal mért vizsgálati eredményét hasonlítjuk össze.

A továbbiakban a változók kvantitatív, ordinális, nominális jellege alapján tekintjük át a szükséges R parancsokat, a lehetséges szignifikancia értékeket, s az eredmények interpretációjának, publikációjának főbb lehetőségeit.

### **Kvantitatív változók az önkontrollos hatásvizsgálatban**

Előfordulhat, hogy az önkontrollos hatásvizsgálat elő- és utótesztbeli változói (a továbbiakban röviden:  $X1$  és  $X2$  változók) kvantitatív skálával jellemezhetők. Ilyen esetben a megfelelő statisztikai próba kiválasztása érdekében arra a kérdésre kell választ találnunk, hogy: az  $X2-X1$  (tehát a két változó különbsége, a továbbiakban röviden:  $X21$ ) normális eloszlású változó-e? Ennek eldöntéséhez használható egyik statisztikai próba a Shapiro-Wilk próba, melynek nullhipotézise: „az  $X21$  különbségváltozó normális eloszlású”; az R parancs pedig a következő:

```
X21 <- d$X2-d$X1
shapiro.test(X21)
```

A Shapiro-Wilk próba elvégzése során az R többek között megadja a szignifikanciát jelző  $p$ -értéket. További teendőinket pedig az határozza meg, hogy ez a  $p$  érték mekkora:

1) Ha a Shapiro-Wilk próba eredménye szerint  $p > 0,05$ , akkor ez azt jelenti, hogy az  $X21$  különbségváltozó normális eloszlású. Ilyen esetben páros t-próbával hasonlíthatjuk össze az elő- és utóvizsgálati eredményeket. A páros t-próba arra a kérdésre ad választ, hogy „ $X1$  és  $X2$  kvantitatív változók átlagai egyenlők-e egy-mással?”; nullhipotézise pedig: az  $X1$  és  $X2$  változó populáció átlaga egyenlő. A páros t-próba R-parancsa most a következő lesz:

```
t.test(d$X1,d$X2, paired=T)
atlagX1=mean(d$X1)
atlagX2=mean(d$X2)
atlagX21=atlagX2-atlagX1
```

Eddigi tevékenységünket pedig a következő mondón foglalhatjuk össze például (a *p-value* helyére az R által megadott *p*-értéket kell beírni): „Vizsgálatunkban arra voltunk kíváncsiak, hogy az IPOO – minimum program tanulás módszertani tréning előtt és után mért tanulás módszertani teljesítményben (a Jupiterbolha-próba összpontszámában) jelentős különbség tapasztalható-e? Különösen: kedvező (itt: teljesítményt növelő) hatása volt-e a tréningnek, vagy sem? Az adekvát statisztikai próba kiválasztása érdekében a két mérés különbségének normális eloszlását Shapiro-Wilk próbával ellenőriztük, amelynek eredménye szerint ( $p=p\text{-value}$ ), a két változóból képzett különbség-változót normális eloszlásúnak tekinthetjük. Ennek ismeretében páros t-próbával ellenőriztük, hogy a Jupiterbolha-próba átlaga lényeges változást mutat-e a két időpont között. Ennek eredménye szerint...”

A szöveg folytatása most a páros t-próba szignifikancia szintjétől függ:

- Ha a páros t-próba eredménye:  $p > 0,05$ , akkor az  $X1$  és  $X2$  változók populáció átlagaira vonatkozó nullhipotézist megtarthatjuk, nincs szignifikáns eltérés a két változó átlaga között. Publikációnkban ezt a bekezdést pedig így fejezhetjük be: „...a Jupiterbolha-próba összpontszáma nem különbözik lényegesen a tréning előtti és utáni mérések alkalmával ( $p=p\text{-value}$ ). Lényegesebb változást nem eredményezett a tréning.”
- Ha a páros t-próba eredménye:  $p \leq 0,05$ , akkor az  $X1$  és  $X2$  változók populáció átlagaira vonatkozó nullhipotézist elvetjük, mert szignifikáns eltérést tapasztaltunk. Korábban megkezdett mondatunkat pedig így folytathatjuk például (a *p-value*, *atlagX1* stb. helyébe az R által megadott számokat kell beírni): „...az utótesztbeli teljesítmény jelentősen eltér az előtesztben nyújtott teljesítménytől (előteszt átlaga = *atlagX1*; utóteszt átlaga = *atlagX2*; átlagok különbsége = *atlagX21*;  $p=p\text{-value}$ ).”

Ha az utótesztbeli értékek jobbak, mint az előtesztbeli értékek, akkor a fejlesztőprogramnak kedvező hatása volt. Ha az utótesztbeli értékek rosszabbak, mint az előtesztbeli értékek, akkor a tréningnek kedvezőtlen hatása volt.

2) Ha a Shapiro-Wilk próba eredménye:  $p \leq 0,05$ , akkor az  $X21$  különbségváltozó nem tekinthető normális eloszlásúnak. Ez azzal a következménnyel jár, hogy az elő- és utóvizsgálat eredményeinek összehasonlításakor nem az átlagokat, hanem a mediánokat vethetjük össze páros Wilcoxon-próbával.

Mindennek szöveges összefoglalása például: „Vizsgálatunkban arra voltunk kíváncsiak, hogy az IPOO – minimum program tanulás módszertani tréning előtt és után mért tanulás módszertani teljesítményben (a Jupiterbolha-próba összpontszámában) jelentős különbség tapasztalható-e? Különösen: kedvező (itt: teljesítményt növelő) hatása volt-e a tréningnek, vagy sem? Az adekvát statisztikai próba kiválasztása érdekében a két mérés különbségének normális eloszlását Shapiro-Wilk próbával ellenőriztük, amelynek eredménye szerint ( $p=p\text{-value}$ ), a két változóból képzett különbségváltozót nem tekinthetjük normális eloszlásúnak. A továbbiakban páros Wilcoxon-próbát végeztünk, amelynek eredménye szerint...”

A páros Wilcoxon-próbával kapcsolatos információkat az ordinális változókkal foglalkozó következő fejezetben foglaljuk össze.

### Ordinális változók az önkontrollós hatásvizsgálatban

Előfordulhat, hogy vizsgált változóink: a) eleve ordinális skálájúak (pl.: „Vizsgálatunkban arra voltunk kíváncsiak, hogy igaz-e, hogy a csoport félév végi matematika felmérő dolgozatra kapott érdemjegyei lényeges változást mutatnak a félév eleji dolgozathoz képest? A páros Wilcoxon-próba eredménye szerint...”) ; b) kvantitatív skálájúak ugyan, ám különbség-változók nem normális eloszlásúak, így csak az ordinális változókkal végrehajtható statisztikai próbákat alkalmazhatjuk esetükben (lásd fentebb a Shapiro-Wilk próba  $p \leq 0,05$  esetében írtakat például).

A statisztikai próbával eldöntendő kérdés most az, hogy „az  $X1$  és  $X2$  (legalább) ordinális változók különbsége nullára szimmetrikus változó-e?”, ahol példánknál maradván  $X1$ =félév elején szerzett érdemjegy;  $X2$ =félév végén szerzett érdemjegy. Megjegyzés: ha  $X1$  és  $X2$  különbsége szimmetrikus, akkor a minta egésze nem mutat változást, hiszen az egyéni szinten jelentkező változások kiegyenlítik egymást. E kérdés megválaszolására alkalmas statisztikai próba a páros Wilcoxon-próba, melynek nullhipotézise: az  $X1$  és  $X2$  változó különbsége szimmetrikus. Az R parancs:

```
wilcox.test(d$X1, d$X2, paired=T)
```

A további teendőinket pedig most is a szignifikancia értéke fogja meghatározni:

1) Ha a páros Wilcoxon-próba eredménye:  $p > 0,05$ , akkor megállapíthatjuk, hogy nincs szignifikáns különbség a két minta között. Az imént félbehagyott mondatot pedig így folytathatjuk: „...a két vizsgálat eredményei között nincs szignifikáns eltérés ( $p=p\text{-value}$ ). Az ún. előjel-próba a szimmetria sérülésének más aspektusát mutathatja ki, mint az imént végzett páros Wilcoxon-próba, ezért célszerűnek láttuk az előjel-próbát is elvégezni, hátha az kimutat valamilyen eltérést.”

A szövegben jelzett előjel-próbához tartozó R parancs a következő:

```
medN1 <- sum(d$X2-d$X1 > 0)
medN2 <- sum(d$X1-d$X2 > 0)
binom.test(medN1, medN1+medN2, 0.5)
```

A fenti R parancsban  $medN1$ = azon értékek száma, ahol  $X1 < X2$ , s  $medN2$  = azon értékek száma, ahol  $X1 > X2$ . Az előjel-próba lehetséges eredményei: ha az előjel-próba eredménye:  $p > 0,05$ , akkor a publikációnknak ezt a bekezdését ehhez hasonlóan zárhatjuk: „Az előjel-próba szerint sincs kimutatható különbség: a jobb és a rosszabb eredményt elérők száma nem tér el jelentősen ( $p=p\text{-value}$ ).”

Azonban, ha az előjel-próba eredménye:  $p \leq 0,05$ , akkor ezt állapíthatjuk meg (a  $medN1$ ,  $medN2$  és  $p\text{-value}$  értékek értelemszerű behelyettesítésével): „Az előjel próba szerint a jobb/gyengébb eredményt elérők száma lényegesen eltér ( $p=p\text{-value}$ ),  $medN1$  esetben nő,  $medN2$  esetben csökken.”

2) Ha a páros Wilcoxon-próba eredménye:  $p\text{-value} \leq 0,05$ , akkor szignifikáns különbség van az  $X1$  és  $X2$  változók eloszlása között. Futtassuk le a következő parancsokat:

```
wilcox.test(d$X1, d$X2, paired=T)
wilcox.test(d$X2, d$X1, paired=T)
```

Mindkét parancs eredményeként az R outputjában egy *V-érték* lesz látható a képernyőn. Az első parancs *V-értékét* jelöljük a továbbiakban *VI*-gyel, a másodikat *V2*-vel, s hasonlítsuk össze értékeiket!

Ha *V2* nagyobb *VI*-nél, akkor összességében szignifikáns növekedést tapasztaltunk. Ezt így fogalmazhatjuk meg például (a fentebb közölt „A továbbiakban páros Wilcoxon-próbát végeztünk, amelynek eredménye szerint...” mondat folytatásaként): „...a két vizsgálat eredményei között szignifikáns eltérést találtunk (rangösszeg, ha előteszt értéke > utóteszt értéke = *VI érték*, rangösszeg, ha előteszt értéke < utóteszt értéke = *V2 érték*,  $p=p\text{-value}$ ). Összességében az összpontszámok értéke nőtt.”

Ha *V2* kisebb *VI*-nél, akkor összességében szignifikáns csökkenést tapasztaltunk: „...a két vizsgálat eredményei között szignifikáns csökkenést találtunk (rangösszeg, ha előteszt értéke > utóteszt értéke = *VI érték*, rangösszeg, ha előteszt értéke < utóteszt értéke = *V2 érték*,  $p=p\text{-value}$ ). Összességében az összpontszámok értéke csökkent.”

A fentiekben természetesen *VI*, *V2* és *p-value* értékeket számszerűen kell a szövegbe szerkeszteni. A további értelmezésben pedig segíthet, ha leíró statisztikákat kérünk a két változó különbségéről. Kérhetünk kvartiliseket (melyek a mintát negyedelik), de kérhetünk részletesebb ábrázolást is. A kvartiliseket és pl. a mintát ötödölő értékeket a következőképpen kaphatjuk meg:

```
X21 <- d$X2-d$X1
quantile(X21)
quantile(X21,c(.2,.4,.6,.8))
```

### **Nominális változók az önkontrollos hatásvizsgálatban**

Amennyiben az általunk vizsgált változók egyes értékei megkülönböztethetők ugyan, de nem rangsorolhatók, akkor az önkontrollos hatásvizsgálat céljából megfogalmazható általános értelemben vett kérdésfeltevésünk így hangzik „*X1* és *X2* nominális változó megfigyelt gyakoriságai eltérnek-e lényegesen egymástól?”. *X1* és *X2* most is az elő-, illetve az utóvizsgálat eredményére utal.

Vegyük például a következő esetet: „Egy fejlesztő program elején és végén a Kozéki-féle „Iskolai orientáció” kérdőív révén felmérjük, hogy a diákok tanulási orientációja közül melyik a legdominánsabb kategória: a mélyreható tanulás, a reprodukív tanulás vagy a szervezett tanulás? Vizsgálatunkban arra voltunk kíváncsiak, hogy a tehetséggondozó programban való részvétel megváltoztatta-e a diákok tanulási orientációjának jellegét.”

A megfelelő statisztikai próba kiválasztása változóink lehetséges értékeinek függvénye: kétértékű (dichotóm) változók esetében McNemar-próbát, több értékű változók esetében marginális homogenitás tesztet kell alkalmaznunk:

Dichotóm (kétértékű) változókkal dolgozunk akkor, ha csak két tanulási orientáció kategóriát (mondjuk a mélyreható orientációt és a reprodukív orientációt) veszünk figyelembe. A statisztikai jellegű kérdés ekkor az, hogy az *X1* és *X2* dichotóm (azaz kétértékű) változók gyakorisága hasonló-e. E kérdésben *X1* = az elővizsgálat változója (értékei legyenek: 1=mélyreható tanulás; 2=reprodukív tanulás), *X2* = az utóvizsgálat változója (értékei: 1=mélyreható tanulás; 2=reprodukív tanulás). A kérdés eldöntésére alkalmazható McNemar-próba nullhipotézise szerint az *X1* és *X2* gyakoriságai hasonlóak. A McNemar-próba R parancsa:



```
mcnemar.test(d$X1, d$X2)
```

Ha a McNemar-próba eredménye:  $p > 0,05$ , akkor az  $X1$  és  $X2$  változó gyakoriságai között nincs lényeges különbség, akkor „McNemar-próbával vetettük össze a két időpontban megállapított tanulási orientáció gyakoriságait és nem találtunk lényeges eltérést ( $p=p\text{-value}$ ).” jellegű mondatot fogalmazhatunk meg publikációnkban. Ha a McNemar-próba eredménye:  $p \leq 0,05$  (vagyis: az  $X1$  és  $X2$  változó gyakoriságai között lényeges különbség van), akkor ezt így is megfogalmazhatjuk: „McNemar-próbával vetettük össze a két időpontban megállapított tanulási orientáció gyakoriságait és lényeges eltérést találtunk ( $p=p\text{-value}$ ).” Közzöljük mindkét mérés esetén a változók gyakoriságait a `table(d$X1)` és `table(d$X2)` parancsokkal, hogy a változás lényegét lássuk!

Kettőnél több értékű nominális változók (pl. ha háromféle lehetséges tanulási orientációt veszünk figyelembe: a mélyreható, a reprodukív és a szervezett) esetében is az alapvető kérdésünk egy önkontrollos hatásvizsgálat alkalmával, hogy  $X1$  és  $X2$  változók gyakoriságai hasonló-e. A változók értékészlete azonban itt már több, mint két értéket számlál -  $X1$  elővizsgálati változó és  $X2$  utóvizsgálati változó lehetséges értékei legyenek mondjuk: 1=mélyreható tanulás; 2=reprodukív tanulás; 3 = szervezett tanulás. Az „ $X1$  és  $X2$  gyakoriságai hasonlóak” nullhipotézis tesztelését most a marginális homogenitás teszttel oldhatjuk meg. R parancs (mivel e próba végrehajtásához szükséges függvény az R programon belül a `coin` csomagban található, az `mh_test()` függvény hívása előtt ki kell adnunk a `library(coin)` parancsot):

```
library(coin)
mh_test(table(d$X1, d$X2))
```

Amennyiben a marginális homogenitás teszt eredménye:  $p > 0,05$ , akkor az  $X1$  és  $X2$  változó gyakoriságai között nincs lényeges különbség. Publikációnkban megfogalmazhatjuk, hogy: „Marginális homogenitás teszttel vetettük össze a két időpontban megállapított tanulási orientációk gyakoriságait és nem találtunk lényeges eltérést ( $p=p\text{-value}$ )”. Ellenben, ha a marginális homogenitás teszt eredménye:  $p \leq 0,05$ , az arra utal, hogy az  $X1$  és  $X2$  változó gyakoriságai között lényeges különbség van. Szövegbe foglalva: „Marginális homogenitás teszttel vetettük össze a két időpontban megállapított tanulási orientációk gyakoriságait és lényeges eltérést találtunk ( $p=p\text{-value}$ )”. Közzöljük mindkét mérés esetén a változók gyakoriságait a `table(d$X1)` és `table(d$X2)` parancsokkal hogy a változás lényegét lássuk!

### Zárógondolatok

Jelen tanulmányban az önkontrollos hatásvizsgálatok elemzésekor hat statisztikai próba alkalmazását és R-nyelvbeli parancsait mutatuk be (lásd: 2. ábra):

- a Shapiro-Wilk próbát alkalmaztuk a kvantitatív változók különbségváltozója normális eloszlásának vizsgálatára. E próba szignifikáns ( $p \leq 0,05$ ) értéke esetén nem normális eloszlás jellemzi a vizsgált változónkat (s a továbbiakban csak ordinális vagy nominális változókra vonatkozó statisztikai próbákat alkalmazhatunk), máskülönben normális eloszlásúnak tekinthető kvantitatív változókat használó statisztikai próbák is végezhetőek vele.

- az elő- és utóvizsgálatok (normális eloszlású különbségváltozóval jellemezhető) kvantitatív változóinak összehasonlítására használtuk a páros t-próbát. A páros t-próba  $p \leq 0,05$  értéke jelez szignifikáns különbséget a két vizsgálat között.
- az ordinális változók összehasonlításakor páros Wilcoxon-próbát, illetve páros előjel-próbát alkalmaztunk fentebb. Ha a páros Wilcoxon-próba, illetve a páros előjel-próba szignifikás  $p \leq 0,05$  értékű, akkor szignifikáns különbségről beszélhetünk.
- kétértékű nominális változók összehasonlítására szolgál a McNemar-próba, melynek  $p \leq 0,05$  szignifikancia szintje lényeges különbséget jelez.
- Három vagy több értékű nominális változók elemzéséhez a marginális homogenitás tesztre láthattunk példát. A marginális homogenitás teszt  $p \leq 0,05$  értéke jelzi a szignifikáns különbséget.

Lényeges ugyanakkor, hogy más statisztikai próbák, módszerek (lásd: Varga, 2000) és vizsgálati elrendezések (Eccles és tsai, 2003) is léteznek hasonló kérdések eldöntésére.

Fontosnak tartjuk megjegyezni, hogy az önkontrollos hatásvizsgálat matematikai statisztikai elemzése nem egy vizsgálati tevékenység végének, mint inkább közbülső állomásának tekinthető. A vizsgálati eredmények átgondolásakor, közlésekor célszerű kitérni arra is, hogy mások eredményei alátámasztják-e az aktuális vizsgálati eredményeinket; mi lehet az eredmények közötti hasonlóság/különbség oka; milyen gyakorlati következményei lehetnek eredményeinknek; milyen korlátai (pl. mintavétel problémája, mérőeszközök esetleges megbízhatatlansága stb.) lehetnek vizsgálatunknak, s milyen további kutatási kérdések és lehetőségek merültek fel a vizsgálat kapcsán.

Amennyiben egy fejlesztő program létjogosultságát önkontrollos hatásvizsgálattal szeretnénk ellenőrizni, akkor az elő- és az utóvizsgálat eredményei között megnyilvánuló olyan szignifikáns ( $p \leq 0,05$ ) különbség létezésére szükséges rákérdeznünk, amelyek:

a) a program céljának megfelelő irányúak (például az IQ tekintetében az utóvizsgálat magasabb, a szorongás esetében az utóvizsgálat alacsonyabb pontszáma jelzi a kedvezőbb irányú változást);

b) a program céljának megfelelő (például a spontán érést, fejlődést meghaladó) mértékűek.

A spontán változás mértékét egyrészt kritériumszintként értelmezett életkori sztandardokkal történő összevetéssel ellenőrizhetjük (a kritériumorientált vizsgálatokhoz szükséges statisztikai eljárásokkal kapcsolatban lásd: Mező és tsai, 2015). Másrészt elavult és/vagy nem reprezentatív és/vagy számunkra nem hozzáférhető és/vagy nem létező életkori sztandardok hiányában a fejlesztésben nem részesített, de az elő- és utóvizsgálatokban résztvevő kontrollcsoport fejlődéséhez viszonyíthatjuk a fejlesztésbe bevont csoport eredményeit (független minták összehasonlítására alkalmas statisztikai próbákkal). Ugyanakkor az is előfordulhat, hogy nem egyfajta spontán fejlődési/érési rátához, hanem egy fejlesztési tervben célként (nem kontrollcsoportok és nem életkori standardok által) előírt elő-/utóvizsgálati különbséghez (pl. megadott értékű pontszámnövekedéshez) mérhetjük az általunk fejlesztett csoportban kapott különbségeket. Ebben az esetben kritériumként megadott pontszám-különbséghez hasonlíthatjuk a fejlesztett csoportban tapasztalt pontszámbeli különbségeket (Mező és tsai, 2015).

Ha sem életkori sztandard, sem kontrollcsoportbeli eredmények, sem fejlesztési tervben meghatározott különbségértékek nem állnak rendelkezésünkre, akkor csak az elő- és utóvizsgálat közötti különbségről tehetünk megállapításokat a jelen tanulmányban közöltek szerint.

**Irodalom**

- Abari K. (2008): A tehetségdiagnosztika adatkezelésbeli alapjai R környezetben. In Mező F. (szerk.): *Tehetségdiagnosztika*. Kocka Kör & Faculty of Central European Studies, Constantine the Philosopher University in Nitra, Debrecen. pp 105-130.
- Abari Kálmán, Mező Ferenc, Mező Katalin és Máth János (2015): Fejlesztőprogramok hatásvizsgálatát szolgáló adatbázisok szerkezete egy ingyenes statisztikai szoftverben: az R-ben. *Különleges Bánásmód*, I. évf. 2015/2. szám, 37-47. DOI 10.18458/KB.2015.2.37
- Eccles, M., Grimshaw, J., Cambell, M. & Ramsay, C. (2003): Research designs for studies evaluating the effectiveness of change and improvement strategies. *Quality and Safety in Health Care*, 2003, 12, 47-52 doi: 10.1136/qhc.12.1.47
- Falus I. & Ollé J. (s.a.): *Statisztikai módszerek pedagógusok számára*. Okker Kiadói Kft., Budapest.
- Máth J. (2004): Kategórikus változók elemzése (loglineáris modell látens változókkal). *Alkalmazott Pszichológia*, VI/1, pp 57-81. Letöltés: 2015.09.14. Web: [http://psycho.unideb.hu/archivum/kategorikus\\_valtozok.pdf](http://psycho.unideb.hu/archivum/kategorikus_valtozok.pdf)
- Máth János, Mező Ferenc, Abari Kálmán és Mező Katalin (2015): Fejlesztőprogramok hatásvizsgálatának matematikai statisztikai alapfogalmai. *Különleges Bánásmód*, I. évf. 2015/1. szám, 69-77. DOI 10.18458/KB.2015.1.69
- Mező F., Máth J. és Abari K. (2008): A különbségvizsgálatokon alapuló tehetségdiagnosztika matematikai statisztikai alapjai (adatelemzési útmutató). In Mező F. (Szerk.): *Tehetségdiagnosztika*. Kocka Kör & Faculty of Central European Studies, Constantine the Philosopher University in Nitra, Debrecen. pp 131-207.
- Mező Ferenc, Máth János, Abari Kálmán és Mező Katalin (2015): Fejlesztőprogramok egymintás, kritériumorientált hatásvizsgálatának matematikai statisztikai háttere. *Különleges Bánásmód*, I. évf. 2015/3. szám, 69-78. DOI 10.18458/KB.2015.3.69
- Reiczigel J., Harnos A. & Solymosi N. (2007): *Biostatisztika nem statisztikusoknak*. Pars Kft., Nagykovácsi.
- Solymosi N. (2005): *R<...erre, erre...! Internetes R-jegyzet*. Letöltés: 2015.09.14. Web: <http://cran.r-project.org/doc/contrib/Solymosi-Rjegyzet.pdf>
- Vargha A. (2000): *Matematikai statisztika pszichológiai, nyelvészeti és biológiai alkalmazásokkal*. Pólya Kiadó, Budapest.