A FEJLESZTŐPROGRAMOK HATÁSVIZSGÁLATÁT SZOLGÁLÓ ADATBÁZISOK SZERKEZETE EGY INGYENES STATISZTIKAI SZOFTVERBEN: AZ R-BEN

Szerző:

Abari Kálmán
Debreceni Egyetem

Mező Ferenc
Debreceni Egyetem

Mező Katalin
Debreceni Egyetem

Máth János
Debreceni Egyetem

Lektorok:

Demetrovics János
Eötvös Lóránd Tudományegyetem

Nagy Dénes
Nemzetközi Szimmetria Társaság

Varga Imre
Szegedi Tudományegyetem

Koncz István
Professzorok az Európai Magyarországért Egyesület

Első szerző e-mail címe:
abari.kalman@arts.unideb.hu

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

Absztrakt

A (gyógy)pedagógiai fejlesztőprogramok matematikai statisztikai alapokon nyugvó hatásvizsgálatához lényeges: 1) ismerni a minimálisan szükséges matematikai alapfoalmakkal (ezeket lásd: Máth és tsai, 2015), 2) egy megfizethető áru statisztikai szoftver, 3) e szoftverek alkalmazásához szükséges - legalább felhasználói szintű – jártasság is. E két utóbbi szükségletre nyújt praktikus megoldást jelen tanulmány, miközben bemutatja az R nyelvet, s annak lehetőségeit.Az R egy ingyenesen letölthető, matematikai statisztikai számítások végrehajtására is alkalmas szoftver, mely praktikus segédeszköze lehet a különleges bánásmódot igénylők kutatásának. Jelen tanulmány az R-ről szóló rövid tájékoztatót követően a kutatások során keletkezett adatok adatbázissá, adattáblává szervezését foglalja össze a fejlesztőprogramok hatásvizsgálatának három jellegzetes esetében.

Kulcsszavak: fejlesztőprogram, hatásvizsgálat, statisztika

Diszciplinák: matematika, pszichológia, gyógypedagógia, pedagógia

Abstract

STRUCTURES OF DATABASES FOR IMPACT STUDIES OF DEVELOPMENT PROGRAMMES IN A FREE STATISTICAL SOFTWARE: IN THE SOFTWARE 'R'

In order to carry out effectiveness study of the development programmes (e.g. education, children with special needs) based on mathematical statistical methods, the follwing factors are important: 1) you need some basic mathematical definitions (Math et all., 2015), 2) a

statistical software (e.g.: the 'R') at a reasonable price, 3) some experience in working with this software. The 'R' is a free downloadable software (that can be applied in mathematical statistical calculations too) which can be a very useful instrument of research of children who need special treatment. This paper provides a practical solution in connection with the last two factors, and introduces language 'R' and its possibilities.

Keywords: development programme, effectivenes study, statitics

Disciplines: mathematics, psychology, special education, pedagogy

A különleges bánásmódot igénylő tanulók számára kidolgozott (gyógy)pedagógiai, pszichológiai fejlesztőprogramok hatásvizsgálata főként különbségvizsgálatok révén valósulhat meg. E különbségvizsgálatok alkalmával:

a) Egymintás különbségvizsgálatról beszélünk, ha egy fejlesztésbe bevont csoport vizsgálati eredményét vetjük egybe a fejlesztési célként kitűzött célértékkel.
b) Összetartozó mintáról beszélünk akkor, ha egy csoport két (pl. fejlesztést megelőző és követő) vagy több alkalommal (pl. több éven keresztül történő nyomon követéses jelleggel) mért vizsgálati eredményeit hasonlítjuk össze.
c) Független mintás vizsgálatról beszélünk akkor, ha két (pl. a fejlesztésbe bevont vizsgálati, s be nem vont kontrollcsoportot) vagy több (pl. három különböző olvasástanítási módszerrel fejlesztett) csoport eredményeit hasonlítjuk össze.

E hatásvizsgálatok közül az egymintás és az összetartozó mintás esetek szervezési szempontból egyszerűbben megoldhatók, hiszen „csak" a fejlesztésbe bevont tanulókról szükséges adatot gyűjteni, s kontrollcsoport(ok) keresésére, vizsgálatára nincs szükség. Éppen ezért egy fejlesztőprogram hatásvizsgálatával kapcsolatban az egymintás, illetve az összetartozó mintás különbségvizsgálatok matematikai statisztikai alapokon nyugvó elemzése fogalmazható meg olyan minimum követelményként, amely a nem eseti jellegű fejlesztőhatás bizonyításához szükséges (a Magyar Tehetségsegítő Szervezetek Szövetsége által koordinált tehetségpont-hálózatban például az önkontrollos hatásvizsgálatokkal történő bizonyítás igénye jelenik meg – Balogh és Mező, 2010). Természetesen a független mintás vizsgálatok további bizonyítékokkal szolgálnak a programok fejlesztőhatásával kapcsolatban.

Bármely különbségvizsgálatot válasszuk is azonban, a szükséges adatok – például megfigyelés, kísérlet, interjú, tartalomelemzés, kérdőív felvétel, tesztelés révén történő – összegyűjtését követően azokat adatbázisba kell rendeznünk, s ezt követően kezdődhet el a matematikai statisztikai szintű adatelemzés. Ugyanakkor: már a vizsgálat(sorozat) tervezésekor célszerű átgondolni a későbbiekben létrejövő adatbázis szerkezetét, s figyelemmel érdemes lenni a matematikai statisztikai számításokhoz alkalmazott szoftvernek az adatbázisokra, adatbázis-szerkezetekre vonatkozó követelményeire.

A fentiek alapján, jelen tanulmány csak az egymintás és az összetartozó mintás hatásvizsgálatokra fókuszál a továbbiakban, s foglalja össze tömören a matematikai statisztikai elemzéshez használható R szoftverrel kapcsolatos leglényegesebb információkat (független mintás esetekkel kapcsolatban lásd: Mező, Máth és Abari, 2008). E tanulmány feltételezi Máth és tsai (2015) összefoglaló tanulmányában közölt matematikai statisztikai alapfogalmak (pl. változó, skála, eloszlás, populáció, minta, hipotézis vizsgálat, szignifikancia) ismeretét.

Az R nyelv

Az R egy magas szintű programozási nyelv és környezet, melynek legfontosabb felhasználása az adatelemzés és az ahhoz kapcsolódó grafikus megjelenítés. Az R egy szabadon használható, többplatformos (Windows, OS X, Unix és Linux operációs rendszereken futó) programcsomag, amely 1995-ös létrehozása óta egyre közkedveltebb a statisztikai elemzés területén. Népszerűsége nem csupán ingyenességének köszönhető. Az R a statisztika és a programozási nyelvek világából érkező szakemberek nemzetközi együttműködésének a terméke, így a kínált statisztikai eljárások szinte végtelen, dinamikusan növekedő tárházat egy teljes értékű, objektum-orientált programozási nyelv segítségével érhetjük el. Az R kiváló grafikus képességekkel rendelkezik és rendkívül jól dokumentált. Elsajátítása azonban nagyobb kezdeti befektetést igényel, ugyanis az R-rel való kommunikáció parancsok egymás utáni megadásán alapul.

Az R nyelv egy ún. interpretált szkript nyelv, azaz az utasításainkat nem fordíthatjuk le futtatható állománnyá, hanem a végrehajtandó parancsokat az R egyesével értelmezi (ellenőrzi a parancs szabályosságát, ha megfelelőnek találja, rögtön végrehajtja). Az R program legfontosabb része tehát az R-értelmező (interpreter), amely a parancsok végrehajtásáért felelős. Az R-interpreterhez kétféle módon juttathatunk parancsot, vagy interaktívan, az egyes parancsok begépelésével, vagy „kötegelten" ún. szkript módban, előre összegyűjtött parancsok formájában.

Az R hivatalos oldala a *http://www.r-project.org*, ahol számos információt találunk az R-környezetről, valamint itt találhatjuk meg az operációs rendszerünknek megfelelő telepítőt is. A Windows verzió közvetlenül a *http://cran.r-project.org/bin/windows/base* oldalról tölthető le. Itt az *R-3.2.2-win.exe* linken kattintva tölthetjük le számítógépünkre az R telepítőt, ugyanis jelen tanulmány írásakor a legfrissebb változat a 3.2.2 volt. Ez a hivatkozás folyamatosan változik, mivel évente kb. 2 új R verzió megjelenése várható. A telepítő program futtatásával az R-t számítógépünkre installálhatjuk.

Az R Windows verziójának indítása után egy rendkívül egyszerű felhasználói felületet kapunk (*RGui*), amelynek legfontosabb eleme az *R Console* nevű ablak. A konzolban megjelenő '>' (nagyobb jel) jelzi számunkra, hogy az R várja a parancsainkat. Ez az ún. prompt, amely után egy tetszőleges karaktersorozat begépelésére van lehetőségünk. A begépelt parancsot az ENTER billentyűvel küldjük el az R értelmezőnek, amely ha mindent „rendben" talál a parancsunkban, akkor végrehajtja azt.

Az alábbiakban a teljesség igénye nélkül összefoglaljuk az R adatkezelésével kapcsolatos fontosabb függvényeket és operátorokat:

1) Az R-környezet parancsai:
- Objektumok listázása: ls(), objects()
- Objektumok törlése: rm()
- Beépített adatobjektumok kezelése: data()
- Csomag letöltése, telepítése: install.packages()
- Telepített csomag betöltése: library()
- Munkakönyvtár kezelése: getwd(), setwd(), dir()
- Szkript futtatása: source()
- Környezetek kezelése: search(), attach(), detach()
- Segítség kérése: help(), ?, help.search(), help.start(), find(), apropos(), example(), vignette(), demo(), RSiteSearch()

2) Alapműveletek parancsai:
- Értékadás: <-, =, <<-, ->, ->>
- Matematikai műveletek: +, -, *, /, ^, **, %/%, %%

- Logikai műveletek: ==,!=, <, >, <=, >=, &, |, !
- Karakteres műveletek: paste(), nchar(), substr()
- Vektor vagy (rendezett) faktor indexelése: x[4], x[2:3], x[-2], x[-(3:5)], x[c(2,4,6)], x[”elemnev”], x[x<5], x[x>7 & x<11]
- Lista vagy adattábla indexelése: d[3], d[[3]], d$elemnev
- Mátrix vagy adattábla indexelése: m[2,3], m[,2], m[,3], m[1:3,-4], m[c(3,2,6), c(2,4)]
- Adatobjektum módja és hossza: mode(), length()

3) Adatok beolvasása és kiírása:
- Adattábla beolvasása tagolt szöveges állományból: read.table(), read.csv(), read.csv2(), read.delim(), read.delim2()
- Fix oszlopszéles állomány beolvasása: read.fwf()
- Adattábla és mátrix kiírása tagolt szöveges állományba: write.table(), write.csv(), write.csv2()
- Egyéb input is output lehetőségek szöveges és bináris állományokkal: dget(), dput(), cat(), load(), save(), save.image()
- Adatok begépelése: scan(), fix(), edit()
- Olvasás a vágóasztalról: readClipboard()

4) Adatobjektumok létrehozása:
- Vektor létrehozása: c(), vector(), : (kettőspont operátor), seq(), rep(), sequence()
- Lista létrehozása: list(), c()
- Faktor létrehozása: factor(), gl(), ordered()
- Mátrix létrehozása: matrix(), rbind(), cbind()
- Adattábla létrehozása: data.frame(), rbind(), cbind()

5) Haladó adatkezelés az R-ben:
- Adatobjektum szerkezetének listázása: str(), ls.str()
- Adatobjektum „típusának” meghatározása: class(), typeof(), is.numeric(), is.data.frame(), is.factor()
- Típuskonverzió: as.numeric(), as.character(), as.logical(), factor()
- Elemnevek, oszlopnevek, sornevek, faktorszintek megváltoztatása: names(), colnames(), rownames(), levels()
- Numerikus változóból faktor: cut(), ifelse()
- Adattáblák egyesítése: merge()
- Adatszerkezet átalakítása: reshape(), stack(), unstack()
- Két faktor értékeinek kombinálása: interaction()
- Kereszttáblák létrehozása: table(), xtabs()

6) Matematikai és statisztikai alapfüggvények:
- Logaritmus, exponenciális, négyzetgyök és abszolút érték függvények: log(x), log10(x), log2(x), exp(x), sqrt(x), abs(x)
- Trigonometrikus függvények: sin(x), cos(x), tan(x), asin(x), acos(x), atan(x)
- Legkisebb és legnagyobb érték: min(x), max(x), range(x)
- Átlag, medián, kvantilisek: mean(x), median(x), quantile(x)
- Szórás, variancia, korreláció: sd(x), var(x), cor(x, y)
- Több leíró statisztikai mutató: summary(d)
- Eloszlásokkal kapcsolatos függvények: dnorm(), pnorm(), qnorm(), rnorm(), pt(), pf(), pbinom()

Magyar nyelven az R programcsomagot részletesebben ismerteti: Reiczigel és mtsai, 2007; Solymosi, 2005; Abari, 2008. Az R-rel végezhető különbségvizsgálatok matematikai statisztikai útmutatóját közli: Mező, Máth és Abari, 2008.

Egymintás hatásvizsgálatokhoz szükséges adatbázis szerkezete az R-ben

Amennyiben egyetlen (fejlesztésbe bevont) csoport vizsgálati eredményeit hasonlítjuk egy fejlesztési célként megjelölt célértékhez (vö.: Mező, Máth és Abari, 2008, 1.1. fejezet), akkor az elemzés alapját képező adatbázist egy legalább két oszlopból álló táblázatként vizualizálhatjuk. Az első oszlop a vizsgálati személyek azonosítására szolgál (név-, jelige- vagy kódszám szerint), a második a vizsgált (pl. *X*) változó adott személyhez tartozó értékeit tartalmazza. E táblázatnak a fejlécen túl a vizsgálati személyek számának megfelelő sora van (1. ábra).

1. ábra: példa az X vizsgálati változóval kapcsolatos egymintás különbségvizsgálat adatbázisának szerkezetére (forrás: a Szerzők)

Általános szerkezet:

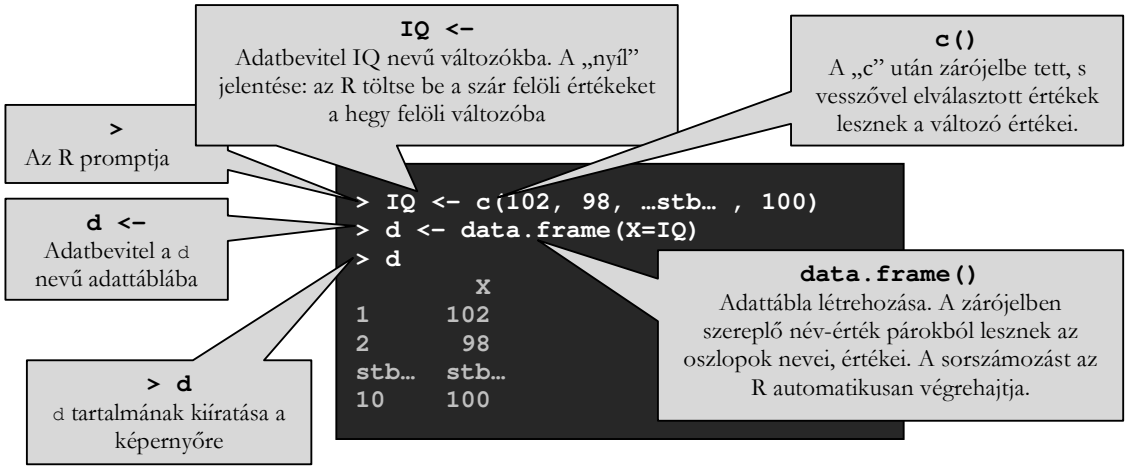
Személy	**X**
1	4
2	7
stb…	stb…
10	8

Konkrét példa:

Tanuló	**IQ**
1	102
2	98
stb…	stb…
10	100

Az 1. ábrán látható adatbázis két formában vonható be az R-rel történő további statisztikai vizsgálatokba: a) egy külső táblázatkezelő vagy szöveges fájlba rögzítjük az adatokat, s ezt a fájlt nyitjuk meg az R számára (lásd a tanulmány későbbi részét); b) közvetlenül az R-be írjuk be az adatokat a 2. ábrán látható módon.

2. ábra: R adatbevitel egymintás különbségvizsgálat esetében egy d nevű adatbázisba (forrás: a Szerzők)



A változók skáláinak tulajdonságai (pl. kvantitatív, ordinális, nominális jellegük, eloszlásuk) függvényében az egymintás különbségvizsgálatokban alkalmazható matematikai statisztikai próbák és R parancsuk például (részletesebben lásd: Mező, Máth és Abari, 2008): Shapiro-Wilk-próba, R parancs: shapiro.test(); egymintás t-próba, R parancs: t.test(); előjel-próba, R parancs: binom.test(); khi-négyzet-próba (χ^2-próba), R parancs: chisq.test().

Két változós összetartozó mintás
hatásvizsgálatokhoz szükséges adatbázis szerkezete az R-ben

Egy fejlesztőprogram önkontrollos hatásvizsgálatának sémája: elővizsgálat → fejlesztés → utóvizsgálat (Mező, Máth és Abari, 2008, 1.2. fejezet). A vizsgálati személyeket és eredményeket rögzítő adattábla ennek megfelelően legalább három oszlopból áll. Ezek közül praktikusan az első oszlop a személyazonosítást szolgálja, a második oszlop az elővizsgálat változóját (pl. *X1* változót), a harmadik oszlop az utóvizsgálat változóját (pl. *X2*) tartalmazza. A sorokat tekintve itt is igaz, hogy a fejlécen túl a vizsgálati személyek számának megfelelő sorból álló táblázatról van szó (3. ábra – R-be történő adatbevitellel kapcsolatban lásd: 4. ábra).

3. ábra: példa az X1 elővizsgálati és X2 utóvizsgálati változóval kapcsolatos összetartozó mintás különbségvizsgálat adatbázisának szerkezetére (forrás: a Szerzők)

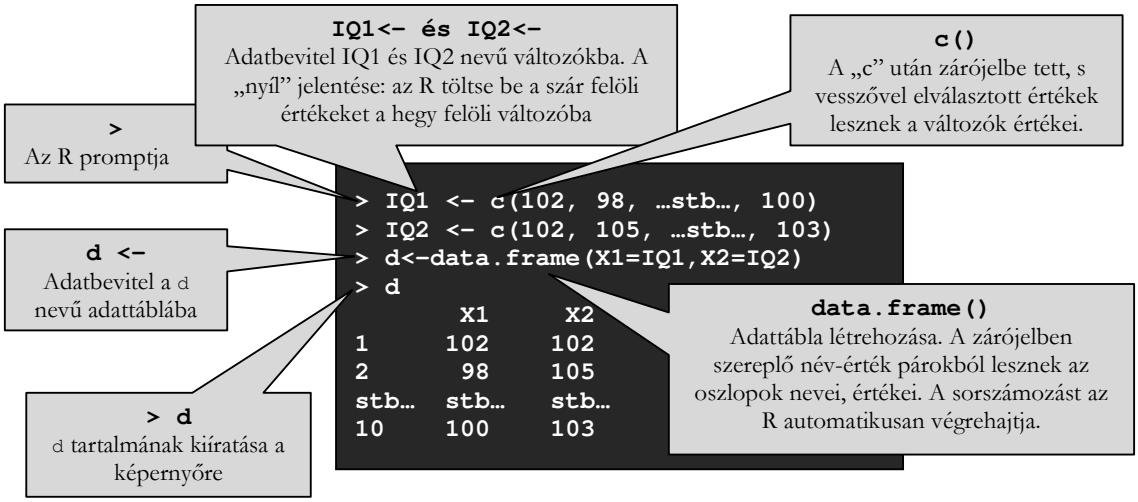
Általános szerkezet:

Személy	X1	X2
1	4	9
2	7	9
stb…	stb…	stb…
10	8	11

Konkrét példa:

Tanuló	IQ1	IQ2
1	102	102
2	98	105
stb…	stb…	stb…
10	100	103

4. ábra: R adatbevitel elő- és utóvizsgálatot magába foglaló összetartozó mintás különbségvizsgálat esetében egy d nevű adatbázisba (forrás: a Szerzők)



Az R számára ebben az esetben is kétféle módon tehetjük elérhetővé adatbázisunkat. Az egyik lehetőség, hogy külső fájlból olvastatjuk be az adatokat – erről a lehetőségről a tanulmány végén részletesebben szólunk még. A másik lehetőség az, hogy magába az R-be írjuk be adatainkat – lásd: 4. ábra.

A változók skálatulajdonságai (pl. kvantitatív, ordinális, nominális jellegük, eloszlásuk) függvényében az elő-/utóvizsgálat jellegű különbségvizsgálatokban alkalmazható statisztikai próbák és R parancsaik például (részletek: Mező, Máth és Abari, 2008): Shapiro-Wilk-próba, R parancs: shapiro.test(); páros t-próba, R parancs: t.test(); páros Wilcoxon-próba, R parancs: wilcox.test(); páros előjel-próba, R parancs: binom.test(); McNemar-próba, R parancs: mcnemar.test(); marginális homogenitás vizsgálat, R parancs: mh_test().

Három vagy több változós összetartozó mintás hatásvizsgálatokhoz szükséges adatbázis szerkezete az R-ben

Egy fejlesztőprogram nyomon követéses jellegű hatásvizsgálatában akár több éven keresztül is történhet adatfelvétel a vizsgálati személyekkel (Mező, Máth és Abari, 2008, 1.3. fejezet). Ez azzal jár, hogy egy-egy vizsgálati személyt három vagy több alkalommal is vizsgálhatunk ugyanazon vizsgálati változó aspektusából, s ez az előzőekhez képest más adatkezelési eljárást fog kívánni tőlünk.

Amennyiben táblázatba rendezzük adatainkat, az esetek egy részében elég lehet a személyek azonosítását szolgáló oszlopot követően annyi további (pl. *X1*, *X2*...*Xn* nevű) oszlopot felhasználni, ahány vizsgálatot végeztünk egy-egy személlyel (5. ábra).

5. ábra: példa az X1, X2 és X3 vizsgálati változóval kapcsolatos összetartozó mintás különbségvizsgálat adatbázisának szerkezetére (forrás: a Szerzők)

Általános szerkezet:

Személy	**X1**	**X2**	**X3**
1	4	9	10
2	7	9	13
stb…	stb…	stb…	
10	8	11	7

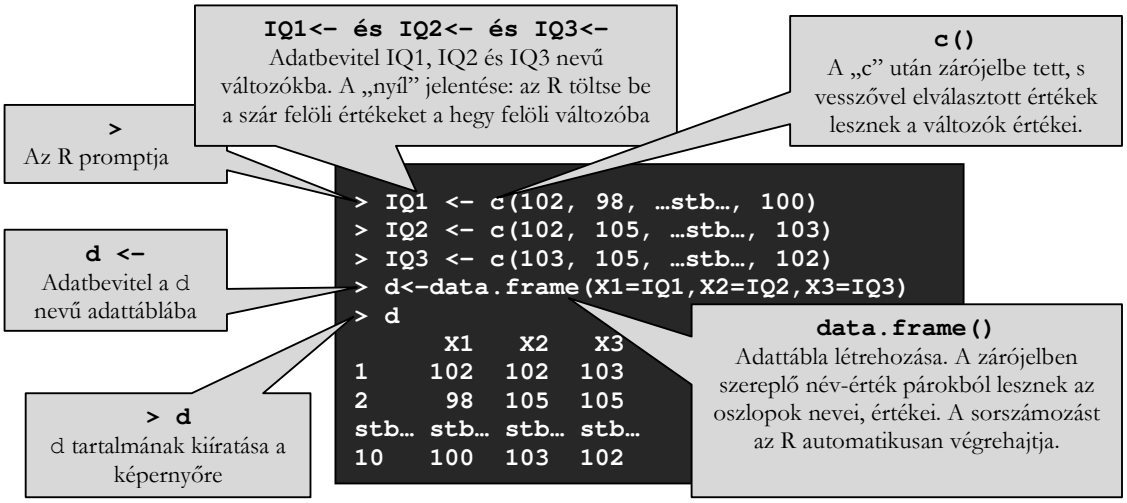
Konkrét példa:

Tanuló	**IQ1**	**IQ2**	**IQ3**
1	102	102	103
2	98	105	105
stb…	stb…	stb…	stb…
10	100	103	102

Az adatokat ebben az esetben is vagy egy külső fájlból hívhatjuk meg, vagy pedig a 6. ábrán látható módon közvetlenül R-ben is létrehozhatunk adattáblát.

Azonban egyes statisztikai próbák az 5. ábrától eltérő (tehát az „egy személyhez tartozó adatok azonos sorba kerüljenek” szabálytól különböző) adatstruktúrát követelhetnek meg. Esetenként arra lehet szükség, hogy az egy személyhez tartozó *n* számú mérés *n* számú külön sorban szerepeljen (lásd: 7. ábra). Ekkor az az elv érvényesül, hogy „az azonos változóhoz tartozó adatértékek kerüljenek azonos oszlopba”. Ebben az esetben *hosszú* adattáblákról beszélünk, míg az előző eset *széles* adattáblákat eredményez.

6. ábra: R adatbevitel három (X1, X2 és X3) vizsgálatot magába foglaló (nyomon követéses) összetartozó mintás különbségvizsgálat esetében egy d nevű adatbázisba (forrás: a Szerzők)



7. ábra: széles és hosszú szerkezetű adattáblák (forrás: a Szerzők)

d adattábla (széles szerkezet):

Személy	X1	X2	X3
1	102	102	103
2	98	105	105
stb…	stb…	stb…	
10	100	103	102

Az adattábla átalakítást az R-ben a reshape() függvény segítségével lehet megoldani.

➡

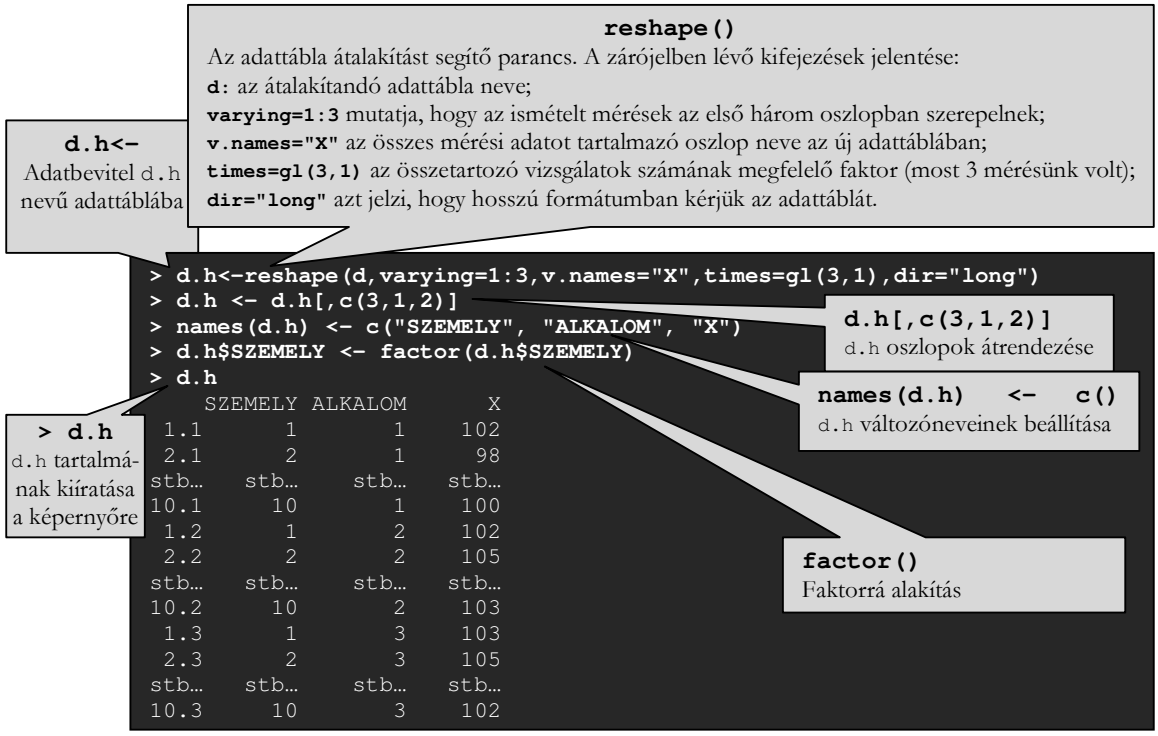
d.h adattábla (hosszú szerkezet):

	SZEMELY	ALKALOM	X
1.1	1	1	102
2.1	2	1	98
stb…	stb…	stb…	stb…
10.1	10	1	100
1.2	1	2	102
2.2	2	2	105
stb…	stb…	stb…	stb…
10.2	10	2	103
1.3	1	3	103
2.3	2	3	105
stb…	stb…	stb…	stb…
10.3	10	3	102

Az R-ben lehetőség van arra, hogy a 7. ábrán látható *d* adattábla szerkezetetét néhány parancs kiadásával a szoftver transzformálja át ugyanezen ábra *d.h* adattábla struktúrájába (8. ábra).

A változók skálatulajdonságai (pl. kvantitatív, ordinális, nominális jellegük, eloszlásuk) függvényében az elő-/utóvizsgálat jellegű különbségvizsgálatokban alkalmazható statisztikai próbák és R parancsaik például (részletek: Mező, Máth és Abari, 2008): Shapiro-Wilk-próba, R parancs: shapiro.test(); Bartlett-próba, R parancs: bartlett.test(); összetartozó mintás egyszempontos varianciaanalízis, R parancs: summary(aov()); Friedman-próba, R parancs: friedman.test(); marginális homogenitás vizsgálat, R parancs: mh_test().

8. ábra: adattábla átstrukturálás az R-ben (forrás: a Szerzők)



Adatok beolvasása és kiírása az R-ben

Adatok beolvasása az R-be: a fentiekben a hatásvizsgálatok során összegyűjtött adatainkat közvetlenül az R-ben rendeztük adattáblákká. Egy másik lehetőség az, amikor az adatokat egy táblázatkezelő vagy adatbázis-kezelő program segítségével rögzítjük, majd mentjük el szöveges állományba. Az adatok exportálása (mentése, fájlba írása) során tipikusan az adatokat elválasztó karaktert kell meghatároznunk, illetve az oszlopnevek állományba írásáról kell gondoskodnunk. Például, ha az Excel magyar változatában a „CSV (pontosvesszővel tagolt)" formátumot választjuk, akkor pontosvesszővel tagolt, oszlopneveket is tartalmazó szöveges állományt hozhatunk létre.

Tagolt szöveges állományokat (például egy c:/temp/adat.txt elérési útvonalú és nevű fájlt) a read.table() függvénnyel olvashatunk be az R-be - például:

> d<-read.table("c:/temp/adat.txt", header = T, sep = ";", dec = ",")

Az első argumentum a beolvasandó szöveges állomány neve, ha szükséges az elérési utat is meg kell adnunk. A header paraméterekben gondoskodhatunk az oszlopnevekről, ha TRUE vagy T az argumentum értéke, akkor az állomány első sorát oszlopneveknek tekinti. Az elválasztó karaktert a sep argumentum tartalmazza. Az állomány tagolására tipikusan a szóközt (sep = " "), a tabulátort (sep = "\t"), a pontosvesszőt (sep = ";") vagy a vesszőt (sep = ",") használjuk. A szöveges állományban a tizedesvessző jelölésére használt karaktert a dec argumentum tartalmazza.

A read.table() függvény helyett használhatjuk a read.csv() és a read.csv2() függvényeket is, amelyek csak a paraméterek alapértelmezett értékeiben térnek el az alapfüggvénytől. Ezekben

a függvényekben a header alapértelmezetten TRUE, az elválasztó karakter pedig a vessző ill. a pontosvessző, valamint a tizedesvessző alakja a pont ill. a vessző. Ha tabulátorral tagolt állományt szeretnénk beolvasni, akkor a read.delim(), ill. a read.delim2() függvényeket érdemes használni, mert az elválasztó karakter itt alapértelmezés szerint a tabulátor karakter.

Adatok kiírása az R-ből szöveges fájlba: adattáblák és mátrixok kiírására a write.table() függvényt használhatjuk. Például:

```
> write.table(szam.telj, "c:/temp/szam_telj.txt", row.names=F, sep=";")
```

Az első paraméter a kiírandó objektum neve, a második pedig a kimeneti állomány elérési útja. A row.names és a col.names logikai paraméterek szabályozzák, hogy a sor- és oszlopnevek szerepeljenek-e a kimeneti állományban. Ezek alapértelmezett értéke TRUE vagy T. Példánkban a FALSE vagy F értékkel letiltottuk a sorneveket. A sep argumentum itt is az elválasztó karakter alakját határozza meg. A write.table() paramétereinek alapértelmezett értékén változtat a write.csv() és write.csv2() függvény.

Összefoglalás

Tanulmányunkban azt tekintettük át és ahhoz igyekeztünk módszertani segítséget nyújtani, hogy az R program miként használható fel (gyógy)pedagógiai fejlesztőprogramok egymintás, illetve összetartozó mintás hatásvizsgálatának adatkezeléséhez. Egymintás vizsgálatról beszélünk, ha vizsgálati személyeinktől összegyűjtött adatainkat egy adott kritériumhoz (pl. fejlesztési tervben meghatározott célértékhez) hasonlítjuk. Ezzel szemben összetartozó mintás vizsgálatról van szó, ha a vizsgálati személyeinktől származó legalább két-két adatot (pl. fejlesztés előtti és utáni eredményeket) vetjük össze. Mint azt fent szintén láthattuk, a szükséges matematikai statisztikai próba megválasztása mindig függ a változó skálájától (tekintve, hogy kvantitatív, ordinális vagy nominális skálájúak-e), kvantitatív változók esetében pedig függ még a skála eloszlásától is (tekintve, hogy normális vagy nem normális eloszlást követ-e).

Noha a tanulmány terjedelmi korlátai nem tették lehetővé az R igazán sokoldalú lehetőségeinek bemutatását, az érdeklődők számára némi segítséget jelenthetnek azonban az alábbi segédletek:

- Az R használatát beépített súgórendszer segíti, bármelyik függvénnyel kapcsolatban kérhetünk segítséget a help() vagy a rövidebb ? operátor segítségével - például:
  ```
  > help(t.test)
  > ?t.test
  ```
- További lehetőség az R megismerésére és a segítségkérésre a help.search(), help.start(), find(), apropos(), example(), vignette(), demo() és az RSiteSearch() függvény.
- Az R-rel kapcsolatban jelen tanulmány szándékánál és lehetőségénél mélyebb jellegű betekintést található magyar nyelven Reiczigel és tsai (2007), Solymosi (2005), illetve Abari (2008) műveiben.
- Az R-rel végezhető különbségvizsgálatok matematikai statisztikai útmutatóját közli: Mező, Máth és Abari (2008), illetve a Különleges Bánásmód folyóirat 2015. és 2016. évben megjelenő számaiban közlésre elfogadott több tanulmány is.

A fejlesztőprogramok hatásáról gyűjtött vizsgálati eredmények R adattáblába rendezését (s az adattábla elmentését, illetve beolvasását) követheti a leíró statisztikai, illetve a hipotézisvizsgálatra épülő matematikai statisztikai adatelemzés. A leíró statisztikai elemzés során egy tanulóra/csoportra vonatkozó átlagértékeket, szórásokat, minimum/maximum értékeke, abszolút és relatív gyakoriságokat, esetleg mediánokat vesszük figyelembe. A matematikai statisztikai elemzés ezzel szemben összetettebb, bonyolultabb számításokat,

valószínűségi ítéletet tartalmaz. De: függetlenül attól, hogy leíró vagy matematikai statisztikai analízist végzünk-e, tartsuk szem előtt, hogy bármely statisztikai elemzés csak annyira lehet jó, amennyire a megelőző adatgyűjtési és adatrendezési munkálatok azt lehetővé teszik. Írásunkkal ez utóbbi, adatrendezési feladathoz kívántunk praktikus támogatást nyújtani.

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.

Balogh László és Mező Ferenc (2010): *Tehetségpontok létrehozása, akkreditációja*. MATEHETSZ, Budapest.

Máth J., Mező F., Mező K. és Abari K. (2015): Fejlesztő programok hatásvizsgálatának matematikai statisztikai alapfogalmai. *Különleges Bánásmód,* 2015, I/1. 69-77.

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.

Reiczigel J., Harnos A. és 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