**Pszichológiai statisztikai elemzések a ROP-R szoftver segítségével és szemléltetésük egy kötődéskutatás adataival[[1]](#footnote-1)**

Vargha András

Professor Emeritus, Károli Gáspár Református Egyetem Pszichológiai Intézete, Budapest

E-mail: vargha.andras@kre.hu

Bánsági Péter

Matematikus mérnök, Budajenő

E-mail: [peter@bansagi.hu](mailto:peter@bansagi.hu)

Jantek Gyöngyvér

Klinikai szakpszichológus, Budapest

E-mail: jantek.gyongyver@gmail.com

**Kivonat**

Cikkünkben egy olyan ingyenes, kétnyelvű (magyar és angol) és egyszerűen használható többváltozós statisztikai programcsomagot (ROP-R) mutatunk be, mely jól használható pszichológiai kutatások statisztikai elemzésében. ROP-R a ROPstat programcsomag olyan fejlesztésének tekinthető, mely a többváltozós statisztika három fontos témakörében (regresszióelemzés, főkomponens- és faktoranalízis, illetve klaszteranalízis) teljes körű statisztikai elemzések végrehajtására alkalmas. ROP-R különlegessége, hogy a statisztikai menüjében kiválasztott és beállított elemzésekhez R-scripteket ír, azokat automatikusan lefuttatja, majd a kapott eredményeket táblázatos formában ROP-R-ben megjeleníti, illetve az eredményekhez tartozó fontosabb grafikonokat standard formátumú (jpg vagy pdf) képfájlokban elmenti. ROP-R moduljainak bemutatása után egy kötődéskutatás adatain végzett többváltozós elemzéssor segítségével szemléltetjük ROP-R hasznosságát a pszichológiai kutatásokban.

**Kulcsszavak**: ROP-R, ROPstat, többváltozós statisztikai elemzések

Manapság a pszichológiai kutatások nélkülözhetetlen részei a statisztikai elemzések, amelyek jól megszerkesztett statisztikai szoftverek segítségével végezhetők el. Ilyen például a ROPstat (Vargha, 2016; Vargha, 2020, Melléklet) is, azon kevés magyar fejlesztésű szoftverek egyike, mely nemzetközi elismertséget is szerzett (Vargha, Torma és Bergman, 2015). A ROPstat egy általános és széles spektrumú statisztikai programcsomag, mely három területen nyújt különleges lehetőségeket:

- robusztus eljárások (R),

- ordinális elemzések (O),

- mintázatfeltáró (pattern oriented) módszerek (P).

Sok közülük semmilyen más ismertebb programcsomagban nem érhető el. Minden érdeme mellett a ROPstat hiányossága, hogy kevés többváltozós statisztikai elemzés végrehajtására alkalmas, például nem lehet benne nemlineáris regressziót vagy dimenziócsökkentő eljárásokat (pl. faktoranalízist) végrehajtani.

Egészen a közelmúltig fizetős programcsomagok (pl. SPSS, SAS, Mplus stb.) uralták a statisztikai szoftverpiacot. Úgy 20 évvel ezelőtt azonban, egy nyílt forráskódú, ingyenes R szoftver megjelenésével új szemlélet kezdett meghonosodni a szoftver-fejlesztők és felhasználók körében. Az R kezdetben egy sima programozási nyelvként indult, melyet statisztikusok fejlesztettek ki matematikai számításokra. Mára hatalmas szoftverrendszerré nőtte ki magát, interaktív adatelemző környezettel. Óriási előnye, hogy ingyenes[[2]](#footnote-2), és hogy szinte minden fontos programozási feladatra létezik R-ben egy vagy több megoldás. Az R CRAN nevű archívumában[[3]](#footnote-3) több mint 16 000 szabadon letölthető adatfeldolgozási és egyéb számolási eljárás áll a felhasználók rendelkezésére széleskörű dokumentációval együtt, speciális programok (az R-ben ezek neve *package* vagy *library*) formájában. R-rel kapcsolatban az egyetlen nehézség, hogy a statisztikai elemzésekhez szigorú szintaktikai szabályoknak eleget tevő utasításcsokrokat (ún. scripteket) kell írni, ahol többnyire még a kisbetű-nagybetű megkülönböztetés is fontos.

E nehézség leküzdésére újabban két felhasználóbarát és az R-hez hasonlóan ingyenes statisztikai programcsomagot is kifejlesztettek, ezek a JASP (JASP Team, 2012) és a jamovi (The jamovi project, 2021; Şahin és Aybek, 2019). Közös alapelvük egy olyan ingyenes és nyílt statisztikai platform kifejlesztése, mely könnyen használható, s amely tartalmazza a legújabb és legfontosabb statisztikai módszereket.

Ezt a filozófiát magukévá téve határozták el a ROPstat szoftver fejlesztői (cikkünk első két szerzője), hogy a ROPstat keretét felhasználva kifejlesztik annak ROP-R nevű többváltozós változatát, mely az R-hez hasonlóan ingyenes, s amely a JASP-hoz és a jamovihoz hasonlóan felhasználóbarát keretet biztosít R-programokon alapuló többváltozós statisztikai elemzések futtatásához.

ROP-R első, 1.0-ás verziója 2022 tavaszán készült el, s a <http://www.bansagi.hu/r/index.html> weboldalról tölthető le. ROP-R legfontosabb jellemzői az alábbiak.

1. ROP-R kétnyelvű (magyar és angol) többváltozós statisztikai szoftver, melynek jelenlegi 10 modulja a többváltozós statisztika alábbi három témakörében kínál teljes körű statisztikai elemzéseket:
   1. Regresszióelemzés
   2. Dimenzió-redukció (főkomponens- és faktoranalízis)
   3. Klaszteranalízis.
2. A kiválasztott statisztikai elemzés minden modul esetén egy átlátható, egyszerű ROP-R feladatablakban paraméterezhető és futtatható.
3. Az elemzés elindítása után ROP-R létrehoz egy R számára olvasható adatfájlt és egy megfelelő egy vagy több R-scriptet, amelyeket lefuttatva a kapott R outputot attraktív formára hozza és elhelyezi azt a ROP-R nézőkében.
4. Ezeket a scripteket ROP-R a felhasználó által elérhető szövegfájlokba írja, amelyek hasznosak lehetnek az R szoftvert tanulók számára az R-scriptek megértésében, az R-ben már korábbi tapasztalatokkal rendelkezők számára pedig a ROP-R-belinél komplikáltabb elemzések R-beli elvégzéséhez.

Ismerve JASP és jamovi széles spektrumú kínálatát a statisztikai elemzésekben, felmerül a kérdés, hogy mi szükség van a ROP-R-re? ROP-R előnyei, mint kompetitív alternatív lehetőségé, az alábbiak szerint fogalmazható meg.

* ROP-R az angolon kívül magyar nyelvre is beállítható, amikor is nemcsak a szoftver kommunikációs nyelve, hanem az eredménylista is magyar nyelvű.
* A ROPstatot ismerő és rendszeresen használók örömmel fedezhetik fel, hogy a ROP-R kerete teljesen megegyezik a ROPstat szoftverével, és mindaz végrehajtható benne, ami ROPstatban (pl. fájlkezelés, szerkesztés, transzformációk stb.), kivéve a statisztikai elemzéseket. Mivel adatkezelési lehetőségei is ugyanazok, ROP-R a ROPstatra specifikus \*.msw adatfájlokat mindenféle konverzió nélkül, simán be tudja olvasni.
* ROP-R számos olyan statisztikai elemzés végrehajtására is alkalmas, amelyek sem a JASP, sem a jamovi, sem más közismert menü vezérelt szoftverben közvetlenül nem elérhetők. Ilyen például a mediációs elemzés és a polinomiális regresszióelemzés, a *k*-medoid és a *k*-medián nemhierarchikus klaszterelemzés vagy a modell-alapú klaszteranalízis.

A jelen cikkben először a ROP-R általános jellemzőit és a futásához szükséges technikai feltételeket foglaljuk össze, majd részletesen ismertetjük a tíz többváltozós statisztikai modult. Végül egy kötődéskutatás (Jantek és Vargha, 2016) adatainak felhasználásával szemléltetjük a ROP-R moduljainak használatát.

**1. ROP-R általános jellemzői és a futtatásához szükséges technikai feltételek**

ROP-R tulajdonképpen a fizetős ROPstat szoftver ingyenes többváltozós változata. A ROPstattal kapcsolatos minden információ megtalálható Vargha, Torma és Bergman (2015), Vargha (2016), illetve Vargha (2019, Melléklet) művében, illetve a [www.ropstat.com](http://www.ropstat.com) weboldalon.

1. ROP-R Windows platformon, a ROP-R.exe program segítségével futtatható, s normális működésének feltétele, hogy az R szoftver (speciálisan annak Rcmd.exe programja) installálva legyen. ROP-R.exe a <http://www.bansagi.hu/r/index.html> weboldalról tölthető le, mely sikeres installálás után a „c:\\_vargha\ropstat” mappában lesz elhelyezve. A program innen is futtatható, de gyakori használat esetén célszerű ROP-R számára parancsikont elhelyezni az asztalon vagy a tálcán.
2. Mivel ROP-R a programból futtat R-scripteket, használatához az R szoftvert telepíteni kell, mégpedig ennek R-4.1.3 verzióját (https://www.filehorse.com/download-r-for-windows/old-versions/), s az ennek során felmásolt Rcmd.exe program elérési útját a ROPstat első használata előtt be kell állítani a ROP-R Beállítások/R-path menüpontjában. Standard telepítés során ez az elérési út: c:\Program Files\R\R-4.1.3\bin\i386\Rcmd.exe, de a c:\Program Files\R\R-4.1.3\bin\x64\Rcmd.exe szintén használható. Ez utóbbi 64 bites eljárások esetén gyorsabb futásokat eredményez, mint az előbbi[[4]](#footnote-4). Újabb R verziók esetén (4.2.0 és afölött) ROP-R nem működik megfelelően.
3. A következő lépésben több R package-et kell installálni, a következőképpen.
   1. R-nek van egy RGui.exe nevű keretprogramja (ugyanabban a mappában, ahol Rcmd.exe), ezt kell először is elindítani.
   2. Ha a gépen korábban már használták RGui-t és installáltak benne R-package-et, akkor simán másolja be az alábbi utasításokat egy csomagban az RGui konzoljába (pl. Ctrl-C, Ctrl-V segítségével), majd nyomja meg az Enter billentyűt.

install.packages("cluster", dependencies = TRUE)

install.packages("stats", dependencies = TRUE)

install.packages("jmv", dependencies = TRUE)

install.packages("psych", dependencies = TRUE)

install.packages("olsrr", dependencies = TRUE)

install.packages("GPArotation", dependencies = TRUE)

install.packages("lavaan", dependencies = TRUE)

install.packages("lavaanPlot", dependencies = TRUE)

install.packages("factoextra", dependencies = TRUE)

install.packages("ggplot2", dependencies = TRUE)

install.packages("ClusterR", dependencies = TRUE)

install.packages("Gmedian", dependencies = TRUE)

install.packages("mclust", dependencies = TRUE)

install.packages("rlang", dependencies = TRUE)

install.packages("MBESS", dependencies = TRUE)

install.packages("MASS", dependencies = TRUE)

install.packages("haven", dependencies = TRUE)

Ennek hatására az RGui installálja az utasításokban megadott package-eket (néhány perc kell hozzá), ami után az RGui-ből ki lehet lépni.

* 1. Ha a gépen RGui-ben korábban még nem installáltak R-package-et, akkor a fenti b. alpont installáló utasításai közül először csak az elsőt másolja be az RGui konzoljába, futtassa azt az Enter billentyűvel, fogadja el a felkínált mentési helyet, majd a sikeres installálás után másolja be a többi utasítást is egy csomagban és futtassa őket. Néha RGui nem szereti az új package-ek egy csomagban való installálását. Ilyenkor érdemes a package-eket egyenként installálni (azaz a fenti lista sorait RGui-be egyenként bemásolni és Enter-rel futtatni).

1. Mindezek után ROP-R elindítható, s az adatállományok ugyanúgy olvashatók be benne, mint ROPstatban. A beolvasáskor az alapértelmezett fájltípus a ROPstat msw típusa. Ezen kívül ROP-R elfogad Excel fájlokat[[5]](#footnote-5) (xls vagy xlsx kiterjesztéssel), szövegfájlokat tabulátorral formattálva vagy csv formátumban, valamint SPSS sav és por adatfájlokat.
2. Az adatfájl beolvasása után a ROP-R moduljai a Többváltozós\_elemzések\_R\_segítségével menüpont segítségével érhetők el.
3. Egy modul elindítása során ROP-R mindig elkészít és futtat egy vagy több R scriptet (kijelezve ezt a képernyőn). Ezután az eredmények tetszetős táblázatokba rendezve megtekinthetők a ROP-R nézőkéjében, ahonnan a táblázati formáka megtartásával Excelbe vagy Wordbe átküldhetők, illetve egyszerűen átmásolhatók.
4. ROP-R fontos tulajdonsága, hogy az elemzésekhez elkészített R-scriptek \*.r formátumú (pl. EFA.r, CFA.r, PolReg.r, MBCA.r stb.) szövegfájlokba íródnak, egy speciális mappában (c:\\_vargha\ropstat\aktualis), melyek az elemzések után a felhasználó által elérhetők, s a ROP-R-ből való kilépés után önállóan is futtathatók R-ben (pl. RGui vagy RStudio segítségével).
5. Ha a futtatás során grafikus ábrák is készülnek (pl. mediációs elemzés vagy konfirmatív faktoranalízis során path diagram, vagy hierarchikus klaszteranalízisben dendrogram), a ROP-R ezeket ugyancsak az 5. pontban említett mappában helyezi el, pdf vagy jpg kiterjesztésű fájlokban.

**2. ROP-R statisztikai menürendszere**

Ebben a fejezetben a ROP-R moduljait a statisztikai elemzés típusa szerint bontva három csoportban mutatjuk be: regressziós elemzések, főkomponens- és faktorelemzések (dimenzió redukciók), illetve klaszteranalízisek, amelyek ROP-R-ben a „Többváltozós elemzések R segítségével” menüponttal futtathatók. Megjegyezzük, hogy ROP-R többi menüpontja (Fájl, Szerkesztés, Esetek, Változók, Transzformációk stb.) logikus, és ugyanúgy használható, mint ROPstatban (vö. Vargha, Torma és Bergman, 2015; Vargha, 2016; Vargha, 2020, Melléklet), illetve a legtöbb más statisztikai szoftverben.

**2.1. ROP-R regressziós moduljai**

ROP-R az alábbi három regressziós modult tartalmazza, melyek a jamovi (The jamovi project, 2021; Şahin és Aybek, 2019) *jmv* R-package-ére, valamint az *rlang* R-package-re (*Henry* és *Wickham*, 2022) építenek:

* Hierarchikus regresszió
* Polinomiális regresszió
* Bináris logisztikus regresszió

Hierarchikus regresszió (HierR)

Ebben a modulban minden függő változót a kijelölt független (magyarázó) változók együttesével magyarázunk egy többszörös lineáris regressziós modellben (vö. de Jong, 1999; Tabachnick és Fidell, 2013, 5. fejezet; Vargha, 2019, 1. fejezet). A független változókat egy blokkindex segítségével különböző csoportokba (blokkokba) sorolhatjuk. A blokkindexek 1 és 9 közötti egész számok, az azonos blokkindexű változók ugyanabba a blokkba tartoznak. HierR segítségével mérhető és tesztelhető, hogy a független változók egymás utáni blokkjai mennyire járulnak hozzá a függő változó bejóslásához.

A többszörös lineáris regresszió modelljébe a blokkok egymás után, egyenként lépnek be, így blokkonként bővül kumulatív módon a regressziós modell, erről kapta ez a módszer a *hierarchikus* nevet. Végső regressziós modellként azt a legegyszerűbb modellt fogadjuk el, ami után már nem nő szignifikánsan az *R*2 többszörös korreláció négyzet, a megmagyarázott varianciaarány. Minden belépő blokk esetén az alábbi statisztikai mutatók láthatók az eredménylistán.

* Az addig belépett független változók többszörös korrelációja (*R*) a függő változóval.
* Az *R* többszörös korreláció négyzete (*R*2), a kumulatív regressziós modell megmagyarázott varianciaaránya.
* A korrigált *R*2 érték, mely a belépett változók száma és az elemszám figyelembevételével korrigálja R2 értékét, hogy ezzel jobban becsüljük az elméleti értéket.
* A kumulatív regressziós modell regressziós becslésének standard hibája.
* A kumulatív regressziós modell szignifikanciája (*F* statisztika, szabadságfokok, *p*-érték) mellyel azt nézzük, hogy a modellbeli független változóknak van-e 0-nál erősebb hatása a függő változóra.
* Az adott blokkhoz tartozó *R*2 növekmény (*R*2+) és szignifikanciája (*F* statisztika, szabadságfokok, *p*-érték).

Az eredménylista tartalmazza az összes kijelölt változó alapstatisztikáit, a végső modell regressziós táblázatát, valamint ezen modellre vonatkozóan a független változók multikollinearitás diagnosztikáját. Az ezzel kapcsolatos táblázatban minden független változó esetén az *R*2 érték a többi független változó által megmagyarázott, TOL = 1 − *R*2 (az ún. tolerancia) a többi független változó által meg *nem* magyarázott varianciaarány, VIF (= variance inflating factor) pedig a TOL reciproka: VIF = 1/TOL. A multikollinearitást, vagyis a változók közti túl erős lineáris összefüggést a nagyon magas *R*2 és VIF, illetve a nagyon alacsony TOL értékek (*R*2 > 0.99, VIF > 100, TOL < 0.01) jelzik. A multikollinearitást okozó független változókat ejteni kell az elemzésből, mert jelenlétük esetén megbízhatatlan lesz a regressziós együtthatók becslése és szignifikanciájuk vizsgálata.

Külön lehetőségként mediációs elemzés is végezhető HierR-ben, ha mindössze két blokkot jelölünk ki úgy, hogy az első blokkban 1 vagy 2 változó szerepel. Ebben az esetben indítás után a program rákérdez arra, hogy akarunk-e mediációs elemzést olyan szerepkiosztással, hogy a második blokk változóit – egyenként – mediátor változóknak tekintjük az első blokk 1 vagy 2 változója, illetve a függő változó kapcsolatában. Ha egynél több függő változót jelölünk ki, akkor a program a mediációs elemzést minden függő változóra külön végrehajtja (feltéve, hogy a kijelölt függő változók száma nem haladja meg a 9-et).

Egy másik lehetőség HierR-ben a regressziós elemzés eredményére nagy hatást gyakoroló, ún. influenszer esetek – többnyire outlierek – azonosítása és az ezt mérő Cook-féle D távolság (Marzjarani, 2015) elmentése. Rendszerint a Dmin = 4/n (itt n az elemzésre kijelölt változók mintájában a komplett esetek száma) küszöbnél nagyobb D-értékű személyeket tekintik outliereknek, bár ez nem kőbe vésett szabály. A HierR modul feladatablakában a „Standard influenszer küszöb (4/n) szorzója” rovat alapértelmezés szerinti 1 értékének megváltoztatásával lehet ezt a küszöböt igény szerint módosítani. Ha ugyanitt az „Influenszer változó elmentése” opciót bejelöljük, akkor ROP-R minden személyre kiszámítja és az aktuális msw adatállományhoz illeszti a D/Dmin hányadossal mért, ún. Cook-féle relatív távolságot.

A modul elemzéseinek végrehajtása után a c:\\_vargha\ropstat\aktualis mappában megtaláljuk az elemzésekhez elkészített ideiglenes adatfájlt (tmpdat.txt), a futtatott R-scriptet (HierReg.r), a mediációs elemzésekhez tartozó path diagramo(ka)t (modplot\*.pdf alakú nevekkel), valamint a részletes R eredménylistát (oo.txt).

Polinomiális regresszió (PolR)

Ebben a modulban minden függő változó polinomiális függését megnézzük minden egyes független változótól. A PolR elemzés fő célja, hogy feltárjunk nemlineáris hatásokat a kijelölt független és függő változók között. PolR valójában egy olyan HierR elemzés, ahol a független változók az adott *X* független változó hatványai (*X*1 = *X*, *X*2, *X*3, …) játsszák a blokkok szerepét, az egymás utáni blokkok a független változó növekvő hatványai (minden blokkban egy-egy ilyen hatvány). A nemlineáris összefüggések az 1-nél nagyobb kitevőjű hatványkomponensek szignifikáns hatásaiként azonosíthatók. PolR regressziós egyenletében *X* képviseli a lineáris komponenst, *X*2 a kvadratikusat, *X*3 a harmadfokú komponenst stb. A PolR modulban a beállítható maximális hatványkitevő 2 és 5 közötti érték lehet (alapértelmezés: 4).

Ha több változó együttes eloszlása többdimenziós normális, az maga után vonja, hogy ezen változók csak lineáris típusú kapcsolatban lehetnek egymással. Emiatt ha a változók között nemlineáris kapcsolatok vannak (ami pl. PolR segítségével is kideríthető), akkor együttes eloszlásuk nem követhet többdimenziós normális eloszlást, ami közismerten mindig egycsúcsú (vö. Vargha, 2019, 1.3. ábra). Mivel egycsúcsú eloszlások esetén nem számíthatunk arra, hogy a minta több, egymástól jól szeparálható, homogén csoportra bontható, a feltárt nem lineáris kapcsolatok pl. a klaszteranalízisekben esélyt adhatnak egy elfogadható klaszterstruktúra feltárására (Vargha, Bergman és Takács, 2016; Vargha és Bergman, 2019).

PolR eredménylistája nagyon hasonlít HierR eredménylistájára, ezért ezt most nem részletezzük. Egy PolR elemzés végrehajtása után a c:\\_vargha\ropstat\aktualis mappában megtaláljuk az elemzéshez elkészített ideiglenes adatfájlt (tmpdat.txt), a futtatott R-scriptet (PolReg.r), valamint a részletes R eredménylistát (oo.txt).

Bináris logisztikus regresszió (BLR)

E modulban minden kétértékű függő változót a kijelölt független változók együttese segítségével regresszálunk. A függő változókat, amelyek közt lehetnek nominális skálájú kategoriális változók is, a blokkindex segítségével különböző blokkokba sorolhatjuk. BLR egyben egy nemlineáris diszkriminancia-analízis is, ahol a függő változó két értéke által definiált két csoportot próbáljuk minél jobban elkülöníteni a kijelölt független változók segítségével (vö. Tabachnick és Fidell, 2013, 10. fejezet).

BLR és HierR mint regressziós elemzések nagyon hasonlóak, a különbségek a következők. BLR függő változója csak kétértékű (dichotóm) változó lehet. A független változók lehetnek folytonosak (ez az alapértelmezés) vagy kategoriálisak. Utóbbi beállításnál ROP-R a változó minden talált értékére készít egy bináris, dummy változót[[6]](#footnote-6).

A BLR modelljébe a blokkok egymás után, egyenként lépnek be. Végső modellként azt a legegyszerűbb modellt fogadjuk el, amely után már nem nő szignifikánsan a független változók predikciós hatása, vagyis nem javul szignifikánsan a regressziós modell illeszkedése.

Statisztikai algoritmusát tekintve BLR egy olyan többszörös lineáris regresszió, a függő változó a kijelölt függő változó nagyobbik értékéhez tartozó mintabeli arányának logitja, a független változók pedig az elemzéshez kiválasztott független változók (a kategoriális változókat természetesen a fentebb leírt módon binarizálva). A logit-tal kapcsolatban megjegyezzük, hogy ha *p* egy tetszőleges szám 0 és 1 között (a mintabeli arányok mindig ilyenek), akkor

odds(*p*) = *p*/(1 − *p*) és logit(*p*) = ln(odds(*p*)),

ahol *ln* a természetes (*e* alapú) logaritmus.

BLR eredménylistája nagyon hasonlít HierR eredménylistájára, ahol minden belépő blokk esetén az alábbi statisztikai mutatók láthatók.

* AIC: a kumulatív BLR modellhez tartozó Akaike-féle információs kritérium.
* R2-McF: a kumulatív BLR modellhez tartozó McFadden-féle pszeudo *R*2.
* R2-CS: a kumulatív BLR modellhez tartozó Cox és Snell-féle pszeudo *R*2.
* R2-Nag: a kumulatív BLR modellhez tartozó Nagelkerke-féle pszeudo *R*2.
* A kumulatív BLR modell szignifikanciája (χ2 statisztika, szabadságfok, *p*-érték), mellyel azt nézzük, hogy a modellbeli független változóknak van-e 0-nál erősebb hatása a függő változóra.
* A belépő blokk szignifikanciája (χ2 statisztika, szabadságfok, *p*-érték), mellyel azt nézzük, hogy a belépő blokk szignifikánsan javítja-e a BLR modell illeszkedését.

Az eredménylista tartalmazza az összes kijelölt változó alapstatisztikáit és a végső modell regressziós táblázatát. Ebben a táblázatban megtaláljuk az esélyhányados (odds ratio = exp(B)) értéket is minden *X* prediktorra, mely az odds arányos változását jelzi, ha *X* értéke 1 egységgel megnő. Ha ez 1-nél nagyobb, akkor X növekedésével várhatóan *p*/(1 − *p*) − és akkor egyben *p* értéke is − nő, vagyis nő a függő változó nagyobbik értékének várható előfordulása a kisebbik értékkel szemben, 1-nél kisebb esélyhányados esetén pedig csökken. Kategoriális változók esetén az esélyhányados érték azt mutatja, hogy milyen lesz az odds arányos változása, ha áttérünk a változó legkisebb (referenciának tekintett) értékéről az adott sorban jelzett másik értékre.

Az eredménylista tartalmazza az összes kijelölt változó alapstatisztikáit, a végső modell regressziós táblázatát, valamint ezen modellre vonatkozóan a független változók multikollinearitás diagnosztikáját. Az ezzel kapcsolatos táblázatban minden független változó esetén az *R*2 érték a többi független változó által megmagyarázott, TOL = 1 − *R*2 (az ún. tolerancia) a többi független változó által meg *nem* magyarázott varianciaarány, VIF (= variance inflating factor) pedig a TOL reciproka: VIF = 1/TOL. A multikollinearitást, vagyis a változók közti túl erős lineáris összefüggést a nagyon magas *R*2 és VIF, illetve a nagyon alacsony TOL értékek (*R*2 > 0.99, VIF > 100, TOL < 0.01) jelzik. A multikollinearitást okozó független változókat ejteni kell az elemzésből, mert jelenlétük esetén megbízhatatlan lesz a regressziós együtthatók becslése és szignifikanciájuk vizsgálata.

Mindezeken felül az eredménylista tartalmazza a független változók multikollinearitás diagnosztikáját (TOL és VIF értékek), a végső modellre vonatkozó klasszifikációs táblázatot, valamint a predikciós mutatók összefoglaló táblázatát. A klasszifikációs táblázatból kiolvasható, hogy a BLR modell alapján történő besorolás rendre hány személyt tesz helyesen a saját csoportba, illetve helytelenül a másik csoportba a függő változó két értéke esetén, a predikciós mutatók pedig minden kumulatív modellre megadják az átlagos helyes besorolási arányt (Pontos%), illetve a kisebb és a nagyobb érték helyes besorolásának arányát (Specif% és Szenzitivitás%). Mindezen besorolások erősen függenek az osztályozási/klasszifikációs küszöbaránytól, mely a BLR feladatablakában adható meg (0 és 1 között, az alapértelmezés 0,50).

A BLR modul elemzéseinek végrehajtása után a c:\\_vargha\ropstat\aktualis mappában megtaláljuk az elemzésekhez elkészített ideiglenes adatfájlt (tmpdat.txt), a futtatott R-scriptet (BinLogReg.r), valamint a részletes R eredménylistát (oo.txt).

**2.2. ROP-R dimenzió redukciós moduljai**

ROP-R az alábbi dimenzió redukciós modulokat tartalmazza:

* Főkomponens-analízis
* Feltáró faktoranalízis
* Konfirmatív faktoranalízis

Főkomponens-analízis (FKA)

Ezzel a modullal kvantitatív változók standard főkomponens-analízise végezhető el (Tabachnick és Fidell, 2013, 13. fejezet; Vargha, 2019, 4. fejezet). Az FKA célja olyan új változók (ún. főkomponensek) létrehozása az input változók súlyozott összegeiként, amelyek az input változók összinformációját jól lefedik és számuk jóval kisebb, mint az input változóké. A lefedettség ilyen mértéke a megtartott főkomponensek kumulatív varianciaaránya, melyet úgy kapunk, hogy leosztjuk ezen főkomponensek sajátértékeinek összegét a változók számával. Ha ez a szám a 70%-ot eléri, jó lefedettségről beszélünk. A kapott főkomponsenseket ortogonális (Varimax) vagy ferde (Promax) forgatással (rotációval) illeszthetjük jobban az input változóegyütteshez. A forgatás által értelmezhetőbb új változókhoz jutunk. FKA eredménylistája az alábbi standard táblázatokat tartalmazza.

* Vizsgált változók alapstatisztikái;
* Sajátértékek és a főkomponensek megmagyarázott varianciaarányai;
* Főkomponens-súlyok mátrixa (korrelációk a változók és a főkomponensek között);
* Forgatás utáni faktorsúly-mátrix (ha kérjük a forgatást);
* Forgatott főkomponensek (faktorok) korrelációs mátrixa (ha ferde forgatást kérünk);
* Struktúramátrix, ami az input változók és a forgatott főkomponensek közötti korrelációkat tartalmazza (ha ferde forgatást kérünk).

Alapértelmezés szerint FKA-ban az 1-nél nagyobb sajátértékű főkomponensek lesznek megtartva, de be lehet állítani azt is, hogy a program egy megadott számú főkomponenst tartson meg (és ha elfogadjuk az alapértelmezést, akkor azokat forgassa is). Külön opcióként kérhető a megtartott főkomponensek mentése, vagyis az elemzett adatállományhoz való illesztése. Ha forgatást kérünk, akkor a forgatott főkomponensek[[7]](#footnote-7) lesznek elmentve.

Ebben a modulban lehetőség van arra is (az *MBESS* R package segítségével; vö. *Kelley*, 2007), hogy megvizsgáljuk az input változók együttesének, mint egy egydimenziós additív skála tételeinek (itemeinek) a belső konzisztenciáját. Ha ezt az opciót bejelöljük, akkor a program kiszámítja az input változókra vonatkozóan a Cronbach-alfa és a McDonald-féle omega reliabilitásmutató mintabeli értékét és az elméleti értékre vonatkozó 95%-os konfidencia-intervallumot (vö. T. Kárász és mtsai, 2022). Ezzel kapcsolatban fontos tudnivaló, hogy a fordított megfogalmazású negatív tételeket az FKA végrehajtása előtt át kell fordítani. Ez megtehető a ROP-R-en belül a Transzformációk menüpont segítségével két lépésben: Transzformációk / Egyváltozós függvények / cX menüpont (c = −1 beállítással), majd a X+c menüpont (c = item minimum + item maximum beállítással). Ha például egy X skála tételeit 5-fokú Likert-skálán mérjük, akkor a konverziós transzformáció: T = 6 – X, vagyis itt a −1-gyel való beszorzás után az X+6 transzformációt kell alkalmazni. Ugyanez egyszerűbben is végrehajtható a ROPstat itemanalízis moduljában, ha bejelöljük ott a „Negatív itemek átfordítása az adatállományban (a régi itemek felülírásával)” opciót.

Az FKA modulban lehetőség van a szélsőséges, extrém esetek (outlierek) azonosítására, valamint az eset extremitás változó elmentésére (ugyanúgy, ahogy HierR-ben az influenszer esetek esetében). A mentéshez csupán az „Eset extremitás elmentése” opciót kell bejelölni. Ez a Mahalanobis távolság egy robusztus változatán alapuló elemzés (vö. Leys és mtsai, 2018) a MASS R-package (Venables és Ripley, 2002) segítségével hajtható végre. ROP-R egy beépített küszöbérték segítségével jelöli ki, hogy mely esetek tekintendők outliernek, de ez a küszöb az „Extremitás beépített küszöbének szorzója” rovat segítségével rugalmasan módosítható. Ezen Outli néven elmentett extremitást mérő változó segítségével ezután megbízható többváltozós elemzések (pl. faktor- és klaszteranalízisek) végezhetők, ha azt feltételes csoportosító változónak jelöljük ki.

Egy FKA elemzés végrehajtása után a c:\\_vargha\ropstat\aktualis mappában megtaláljuk az elemzésekhez elkészített ideiglenes adatfájlt (tmpdat.txt), valamint a futtatott R-scripteket (PCA.r, rotate.r).

Feltáró faktoranalízis (EFA)

Ezzel a modullal – a *psych* (Revelle, 2022), *olsrr* (Hebbali, 2020) és *GPArotation* (Bernaards és Jennrich, 2005) R-package segítségével – kvantitatív változók feltáró faktoranalízise (EFA) végezhető el 3 különböző módszerrel: maximum likelihood (ML), főfaktor-elemzés (angolul principal axis factoring vagy PAF) és minimum reziduális módszer (angolul minimum residual vagy MinRes) (lásd Osborn, 2014; Tabachnick és Fidell, 2013, 13. fejezet). EFA fő célja, hogy feltárjunk egy olyan látens faktormodellt, amelyben a látens faktorok kellő mértékben magyarázzák az input változók közös részét. Ezt a közös részt leggyakrabban a Kaiser-Meyer-Olkin (KMO) mutatóval szoktak mérni, amit 0,70 fölött jónak, 0,80 fölött kiválónak tekinthetünk (vö. Vargha, 2019, 5.1. táblázat). Egy-egy input változó többivel közös részét kezdeti kommunalitásnak, a végső feltárt modellben a látens faktorok által megmagyarázott részét pedig végső kommunalitásnak nevezzük. A faktorok által meg nem magyarázott rész a változó unicitása, amiben a mérési hiba is benne van.

EFA-ban az elsődlegesen feltárt ún. kezdeti faktorstruktúrát ortogonális (Varimax) vagy ferde (Promax) forgatással illeszthetjük jobban az input változóegyütteshez. EFA eredménylistája a vizsgált változók alapstatisztikáin kívül az alábbi EFA-ban szokásosan kiszámított statisztikai eredményeket tartalmazza.

* Kaiser-Meyer-Olkin mutató (KMO);
* Sajátértékek és a főkomponensek megmagyarázott varianciaarányai;
* Elsődleges faktorsúlymátrix (korrelációk a változók és a faktorok között, ha nem kérünk forgatást) és faktoronként megmagyarázott varianciák;
* Az input változók multikollinearitás diagnosztikája (R2, TOL, VIF értékek) − a multikollinearitást okozó változókat ki kell hagyni az elemzésből;
* A faktorilleszkedés tesztelése (a szignifikáns eredmény rossz illeszkedésre utal);
* A feltárt faktormodellt jellemző néhány illeszkedési mutató (RMSEA, RMSR, CFI, TLI);
* Mintázatmátrix, más néven forgatott faktorsúlymátrix (ha kérünk forgatást), az input változók unicitásával és kommunalitásával;
* Forgatott látens faktorok becsült korrelációs mátrixa (ha ferde forgatást kérünk);
* Struktúramátrix, ami az input változók és a forgatott látens faktorok közötti korrelációk becslését tartalmazza (ha ferde forgatást kérünk).

Alapértelmezés szerint EFA-ban az 1-nél nagyobb sajátértékű főkomponensek száma dönti el, hogy hány faktoros struktúrát keressen a program, de meg lehet adni erre egy konkrét értéket is (és ha elfogadjuk az alapértelmezést, akkor az elsődlegesen feltárt faktorstruktúrát el is forgatja). Külön opcióként kérhető a feltárt faktorok mentése, vagyis az elemzett adatállományhoz való illesztése. Ha forgatást kérünk, akkor a forgatott faktorok lesznek elmentve.

Megjegyezzük, hogy ortogonális forgatás esetén a faktorsúlymátrix közös becslése egyrészt a standardizált input változók regressziós becslésének a feltárt látens faktorok mint független változók segítségével, másrészt az input változók és a feltárt látens faktorok közti korrelációknak. Ferde forgatás esetén azonban ezek eltérnek, az előbbit mintázatmátrixnak (angolul pattern matrix), az utóbbit (a korrelációk becslését) pedig struktúramátrixnak (angolul structure matrix) nevezzük.

Egy EFA elemzés végrehajtása után a c:\\_vargha\ropstat\aktualis mappában megtaláljuk az elemzésekhez elkészített ideiglenes adatfájlt (tmpdat.txt), valamint a futtatott R-scripteket (Rbegin.r, EFA.r).

Konfirmatív faktoranalízis (CFA)

Ezzel a − *lavaan* (Rosseel, 2012) és *lavaanPlot* (Lishinski, 2021) R-package-re épülő − modullal konfirmatív faktorelemzés (CFA) végezhető, mely által tesztelhető egy hipotetikus (pl. egy másik mintán EFA-val feltárt vagy elméletileg kigondolt) faktorstruktúra elfogadhatósága, illetve jósága (Harrington, 2009; Rosseel és mtársai, 2018; Vargha, 2019, 6. fejezet).

Egy blokkindex segítségével jelölhető ki, hogy a kiválasztott input változók rendre mely faktorokhoz (skálákhoz) sorolandók, ennek értéke 1 és 9 közötti egész szám lehet. Azonos blokkindexű változók azonos faktorhoz tartoznak.

CFA-ban egy faktormodell elfogadhatóságát χ2-próbával tesztelhetjük, jóságát pedig különböző adekvációs mutatók segítségével mérhetjük. A definiált faktormodell becslésére három alapmódszer áll rendelkezésre (ML = maximum likelihood módszer; ULS = Unweighted least squares, vagyis a súlyozatlan legkisebb négyzetek módszere; végül DWLS = Diagonally weighted least squares, vagyis az átlósan súlyozott legkisebb négyzetek módszere), s ezekre építve öt robusztus becslési variáns:

* MLMV: ML becslés robusztus standard hibákkal, valamint korrigált átlagot és varianciát alkalmazó próbastatisztikával;
* MLR: ML becslés Huber-White robusztus standard hibákkal, valamint egy Yuan-Bentler-féle statisztikával aszimptotikusan megegyező tesztstatisztikával;
* ULSMV: az ULS becslés robusztus variánsa robusztus standard hibákkal, valamint korrigált átlagot és varianciát alkalmazó próbastatisztikával;
* WLSM: DWLS becslés robusztus variánsa robusztus standard hibákkal, valamint korrigált átlagot alkalmazó próbastatisztikával;
* WLSMV: DWLS DWLS becslés robusztus variánsa robusztus standard hibákkal, valamint korrigált átlagot és varianciát alkalmazó próbastatisztikával.

CFA futtatása során mindig megkapjuk a kiválasztott robusztus módszer mellett a megfelelő ML, ULS vagy DWLS alapbecsléshez tartozó eredményeket is.

Alapértelmezésben a blokkindexek segítségével definiált faktorstruktúra látens faktorai korrelálhatnak egymással, de az egyes faktort alkotó tételek reziduálisai (faktor által meg nem magyarázott részei) nem − sem az azonos faktorba tartozó tételek, sem a különböző faktorokba tartozó tételek esetén. Ugyanígy alapértelmezésben nincs megengedve kereszttöltés sem, vagyis az, hogy egy tétel korreláljon valamelyik nem saját faktorral. Ez a legegyszerűbb faktorstruktúra nem mindig teljesül, néha meg kell engedni, hogy a fenti kapcsolatokat képviselő kovarianciák 0-tól különböző értékeket is felvehessenek. Erre utaló jelzést a modifikációs indexek magas értékei adhatnak az elsődleges futás eredménylistáján. Ennek alapján CFA paneljén beállítható, hogy a faktorokon belüli vagy különböző faktorokba tartozó tételek közötti, illetve tételek és nem saját faktorok közötti kovarianciák milyen modifikációs küszöb felett legyenek beépítve egy javított faktormodellbe.

CFA eredménylistája a vizsgált változók alapstatisztikáin kívül az alábbi CFA-ban szokásosan kiszámított statisztikai eredményeket tartalmazza.

* A vizsgált faktormodell modifikációs indexei;
* A vizsgált faktormodell és a kijelölt változók között semmilyen kapcsolatot nem feltételező alapmodell tesztelése χ2-próbával;
* A tesztelt modellek jóságát mérő adekvációs mutatók (AIC, BIC, RMSEA, pClose, CFI, TLI, SRMR);
* Az elsőként vizsgált faktormodell standardizált faktorsúlyai és kommunalitásai;
* A látens faktorok páronkénti standardizált kovariancia (azaz korreláció) becslései;
* A kovarianciákkal javított faktormodell legnagyobb modifikációs indexei (ha van olyan modifikációs index, mely az első futás során meghaladja beállított küszöbértékét);
* A javított faktormodell standardizált faktorsúlyai és kommunalitásai kereszttöltésekkel (ha az elsőként vizsgált faktormodellben bármely kereszttöltéshez tartozó modifikációs index meghaladja beállított küszöbértékét);
* A standardizált reziduális kovarianciák (azaz reziduális korrelációk) becslései a javított faktormodellben (ha az elsőként vizsgált faktormodellben bármely reziduális kovarianciához tartozó modifikációs index meghaladja beállított küszöbértékét).

Egy CFA elemzés végrehajtása után a c:\\_vargha\ropstat\aktualis mappában megtaláljuk az elemzésekhez elkészített ideiglenes adatfájlt (tmpdat.txt), a futtatott R-scripteket (CFA.r, CFA2.r), a részletes R-eredménylistákat (oo.txt, o2.txt), valamint a faktormodellek elkészített path diagramjait (pathplot1.pdf and pathplotR1.pdf). Ha feltételes csoportosító változót is használunk, minden csoport elemzése során készül ilyen diagram.

**2.3. ROP-R klaszterező moduljai**

ROP-R a klaszterező eljárások (vö. Bergman, Magnusson és El Khouri, 2003; Kaufman és Rousseeuw, 2009) széles választékát nyújtja, amelyek közül több nem érhető el más menü vezérelt, felhasználó-barát statisztikai szoftverben. ROP-R-ben az alábbi klaszterező modulok érhetők el:

* Agglomeratív/összevonó hierarchikus klaszteranalízis (AHKA)
* Osztódó hierarchikus klaszteranalízis (OHKA)
* *k*-középpontú (*k*-centrumú) klaszteranalízis (KKA)
* Modell-alapú klaszteranalízis (MKA)

AHKA és OHKA eredményeként egyaránt egy hierarchikus klasszifikáció sorozatot kapunk, melyek a kijelölt kvantitatív input változók alapján készülnek, s amelyeket szemléltetésére dendrogramot szoktak használni (Roux, 2018; Vargha, 2022, 5. fejezet). AHKA első lépésében minden személyt egy egyelemű klaszternek tekintünk, majd minden ez utáni lépésben összevonjuk közös klaszterbe az egymáshoz legközelebbi két klasztert (agglomeratív vagy összevonó algoritmus) mindaddig, amíg a teljes mintát tartalmazó egyetlen nagy klaszterhez nem jutunk.

Ezzel szemben OHKA első lépésében a teljes mintát egyetlen nagy klaszternek tekintjük, s megpróbáljuk két részre bontani úgy, hogy a kapott klaszterek a lehető leghomogénebbek legyenek. Minden további lépésben megkeressük a legheterogénebb klasztert, s a fenti módon két alklaszterre bontjuk (osztódó algoritmus) mindaddig, amíg minden klaszter egyelemű nem lesz (ez AHKA kezdő állapota).

KKA egy olyan nemhierarchikus klaszteranalízis, ahol megpróbáljuk a teljes mintát egy előre megadott *k* számú klaszterre úgy szétbontani, hogy ezek a klaszterek a lehető leghomogénebbek legyenek és egymástól is jól elkülönüljenek (Kaufman és Rousseeuw, 2009, 2. fejezet; Vargha, 2022, 6. fejezet).

MKA modelljében az a kiinduló feltételezés, hogy mintánk adatai egy többdimenziós − többnyire normális − keverékeloszlást követnek (mintha valaki összeöntött volna olyan sokaságokat, amelyek mindegyike valamilyen többdimenziós normális eloszlást követ, de más centrumokkal). Ebben a keretben MKA célja, hogy azonosítsa az összekevert többdimenziós eloszlások számát (optimális klaszterszám) és megadja az eloszlások jellemzőit (Fraley és Raftery, 2002; Gergely és Vargha, 2021; Vargha, 2022, 7. fejezet). MKA-ban minden eloszlás centrumát egy klasztercentrumnak tekintjük, s a klaszterek úgy jönnek létre, hogy minden személyt a hozzá legközelebbi centrum által képviselt klaszterhez sorolunk.

Mindezen lehetőségek ellenére ROP-R nem fedi le az összes olyan klaszterelemzéssel kapcsolatos eljárást, amely a mintázatfeltáró elemzésekben gazdag ROPstat szoftver palettáját jellemzi (vö. Vargha, Torma és Bergman, 2015; Vargha, Bergman és Takács, 2016). E tekintetben ROP-R a ROPstat egyfajta bővítésének tekinthető, melyet a ROPstattal együtt érdemes használni.

Agglomeratív hierarchikus klaszteranalízis (AHKA)

Ez a modul összevonó (agglomeratív) hierarchikus klaszteranalízist (Roux, 2018) végez a mintabeli eseteken a kijelölt input változók felhasználásával, hat lehetséges személytávolsággal (négyzetes euklideszi, euklideszi, Manhattan, Canberra, Maximum, Minkowski), valamint nyolc választható klaszterösszevonási módszerrel (átlagos távolság, minimális távolság, maximális távolság, centroid, medián, Ward, flexibilis béta, McQuitty), a *stats* (R Core Team, 2021) és a *cluster* (Maechler és mtsai, 2022) R-package felhasználásával. AHKA-ban négy választható diagram (dendrogram, Silhouette-ábra, összhiba diagram, Banner diagram) segíti az eredmények értelmezését.

A felhasználó megadhatja klaszterszámok egy övezetét, amelyek mindegyikére az alábbi eredmények láthatók.

* A klaszterstruktúra három adekvációs mutatója (HCátlag, EESS% = megmagyarázott varianciaarány, XBmod = módosított Xie-Beni index; vö. Vargha, Bergman és Takács, 2016; Vargha, 2022, 4.4.4. alpont);
* Klaszterstatisztikák (elemszám, átlag, szórás, minimum, maximum);
* Standardizált átlagok mintázata.

Külön kérésre – a klaszterszámok egy megadott övezetére – a klaszterkódot személyenként megadó klaszterváltozók elmenthetők (az adott msw adatfájlhoz illeszthetők).

Egy AHKA elemzés végrehajtása után a c:\\_vargha\ropstat\aktualis mappában megtaláljuk az elemzéshez elkészített ideiglenes adatfájlt (tmpdat.txt), a kért klaszterszámokhoz tartozó klaszterváltozókkal kiegészített ideiglenes adatfájlt (tmpdat2.txt), a futtatott R-scriptet (AHKA.r), valamint a kért diagramokat jpg vagy pdf fájlban (pl. Dendr1.jpg vagy Banner1.pdf). Ha feltételes csoportosító változót is kijelölünk, minden feltételes csoport elemzése során elkészülnek a kért diagramok.

Osztódó hierarchikus klaszteranalízis (OHKA)

Ez a modul DIANA osztódó hierarchikus klaszteranalízist (Kaufman és Rousseeuw, 2009, 6. fejezet; Vargha, 2022, 5.4.4. alpont) végez az eseteken a kijelölt input változók felhasználásával, 6 lehetséges személytávolsággal (négyzetes euklideszi, euklideszi, Manhattan, Canberra, Maximum, Minkowski), a *cluster* (Maechler és mtsai, 2022), *factoextra* (Kassambara és Mundt, 2020) és *ggplot2* (Wickham, 2016) R-package felhasználásával. A választható diagramok, az eredménylista és a klaszterkódváltozók elmentési lehetősége ugyanaz, mint AHKA-ban.

Egy OHKA elemzés végrehajtása után a c:\\_vargha\ropstat\aktualis mappában megtaláljuk az elemzéshez elkészített ideiglenes adatfájlt (tmpdat.txt), a kért klaszterszámokhoz tartozó klaszterváltozókkal kiegészített ideiglenes adatfájlt (tmpdat2.txt), a futtatott R-scriptet (OHKA.r), valamint a kért diagramokat jpg vagy pdf fájlban. Ha feltételes csoportosító változót is kijelölünk, akkor minden feltételes csoport elemzése során elkészülnek a kért diagramok.

*k*-centrumú klaszteranalízis (KKA)

Ezzel a modullal k-centrumú klaszteranalízis végezhető, 3 különböző módszerrel (*k*-közép, *k*-medoid és *k*-medián elemzés). Az optimális klaszterszám meghatározására többféle ábra is készíthető.

A *stats* (R Core Team, 2021), *cluster* (Maechler és mtsai, 2022), *ClusterR* (Mouselimis, 2022), *Gmedian* (Cardot, 2022) *factoextra* (Kassambara és Mundt, 2020) és *ggplot2* (Wickham, 2016) R-package-et felhasználó KKA modulban 3 különböző típusú *k*-centrumú klaszteranalízis végezhető: *k*-közép, *k*-medoid és *k*-medián elemzés. Utóbbi kettőt akkor szokták javasolni, ha az input változók normalitása súlyosan sérül, vagy ha tiszta ordinális változók (Kaufman és Rousseeuw, 2009, 2. fejezet; Vargha, 2022, 6.1.2. és 6.1.3. alpont). A KKA elemzések közül legismertebb *k*-közép elemzés három választható algoritmussal (Hartigan-Wong, MacQueen, Lloyd/Forgy) áll rendelkezésre.

A KKA modulban több ábra (Silhouette, EESS%, átlagos heterogenitás, f(K) torzulás) segít az optimális klaszterszám meghatározását. KKA eredménylistájának szerkezete és a mentési opciók hasonlóak ahhoz, amit AHKA-ban láttunk.

Egy KKA elemzés végrehajtása után a c:\\_vargha\ropstat\aktualis mappában találjuk az elemzéshez elkészített ideiglenes adatfájlt (tmpdat.txt), a megadott klaszterszámhoz tartozó klaszterváltozóval kiegészített ideiglenes adatfájlt (tmpdat2.txt), a futtatott R-scriptet (KCA.r), valamint a kért diagramokat jpg fájlban. Ha feltételes csoportosító változót is kijelölünk, akkor minden feltételes csoport elemzése során elkészülnek a kért diagramok.

Modell-alapú klaszteranalízis (MKA)

Ezzel a modullal modell-alapú klaszteranalízis (MKA) végezhető. Ennek során a program beállítható klaszterszámok (2 és 25 között) és modelltípusok (14 közül bármely részegyüttes, akár mind a 14 kiválasztható) mindegyikére a program megkeresi a legjobb maximum likelihood illesztést (Fraley és Raftery, 2002; Vargha, 2022, 7. fejezet). MKA-ban a legnagyobb BIC vagy ICL értékű klaszterstruktúra modelljét szokták a legjobbnak tekinteni. A modulban kijelölt elemzéseket ROP-R az *mclust* (Scrucca és mtsai, 2016), *factoextra* (Kassambara és Mundt, 2020) és *ggplot2* (Wickham, 2016) R-package segítségével végzi el.

Az eredmények kiértékelése a BIC vagy ICL ábra, valamint az MKA által azonosított optimális megoldás összefoglaló táblázatai alapján történhet. MKA eredménylistája és a mentési lehetőség nagyon hasonlít KKA-éra. MKA-ban opcionálisan kérhető a BIC- és az ICL-értékek táblázata, a klaszterbesorolások *p*-értékei (besorolások valószínűsége minden személyre és minden klaszterre), valamint a bizonytalanság értékek új változóként való elmentése az adatfájlban.

Egy MKA elemzés végrehajtása után a c:\\_vargha\ropstat\aktualis mappában találjuk az elemzéshez elkészített ideiglenes adatfájlt (tmpdat.txt), az optimális MKA megoldás klaszterszámához tartozó klaszterváltozóval kiegészített ideiglenes adatfájlt (tmpdat2.txt), a futtatott R-scriptet (MBCA.r), valamint a kért ábrákat (pl. BIC ábra vagy ICL ábra) jpg fájlban. Ha feltételes csoportosító változót is kijelölünk, akkor minden feltételes csoport elemzése során elkészülnek a kért diagramok.

**3. ROP-R használatának szemléltetése egy kötődéskutatás adataival**

Egy kötődéssel kapcsolatos pszichológiai kutatásban az ECR-RS (Experiences in Close Relationships – Relationship Structures) kérdőív (Fraley és mtársai, 2011) magyar populációra való adaptációja volt a fő feladat egy 336 fős felnőtt magyar mintán (124 férfi és 212 nő), akik mind heteroszexuális kapcsolatban éltek (Jantek és Vargha, 2016). Az ECR-RS a kötődést két dimenzió (elkerülés és szorongás) alskálájával méri négy kötődési személy (romantikus partner, anya, apa, barát) tekintetében. A 40 tételes kérdőív minden viszonylatban ugyanazt a 10 tételt alkalmazza, melyek közül az első hat az elkerülést, az ez utáni 4 tétel pedig a szorongást méri.

Fraley modelljében biztonságos, jó kötődéssel rendelkeznek azok, akik az elkerülés és a szorongás tekintetében egyaránt alacsony szinten vannak és félelemteli, elkerülő kötődésűek a mindkét dimenzión magas értékűek. A magas elkerülés – alacsony szorongás kombináció az elutasító-elkerülő, a magas szorongás – alacsony elkerülés kombináció pedig az elárasztott-megszállott típusra jellemző (vö. Jantek és Vargha, 2016, 1. ábra).

A vizsgálatban az ECR-RS kérdőív mellett felvételre került – többek között – három másik kérdőív is: a Rövidített Személyiségvonás Kérdőív (Big Five Inventory, BFI-44, vö. John, Donahue és Kentle, 1991; John, Naumann és Soto, 2008; magyar adaptáció: Rózsa, Tárnok és Nagy, 2020), az 5-tételes WHO Jóllét Skála (WBI-5 Well-Being Scale; vö. Bech, 1996, 2012; magyar adaptáció: Susánszky, Konkoly Thege, Stauder és Kopp, 2006), illetve a Beck-féle Depresszió Kérdőív (BDI) rövidített változata (Beck és Beck, 1972; magyar adaptáció: Kopp, Skrabski és Czakó, 1990).

Az alábbiakban az ECR-RS adaptációs vizsgálataival kapcsolatban szemléltetjük a ROP-R számos moduljának használatát.

**3.1. Főkomponens-analízis**

Első lépésben FKA elemzést végeztünk mind a négy kötődési személy 10-10 tételén a ROP-R FKA modulja segítségével. Illusztrációként a romantikus partner 10 tételén elvégzett elemzés eredményét részletezzük.

FKA-ban az első három sajátérték haladta meg az 1 értéket (λ1 = 4,79, λ2 = 1,41, λ3 = 1,03), a 10 tétel összvarianciájának 72,4%-át magyarázva. Ez azt jelzi, hogy az első 3 főkomponens elfogadhatóan helyettesíti a 10 tételt. Ha elforgatjuk ezt a 3 főkomponenst (ferde Promax forgatást választva), a 72,4%-os összvariancia nem változik, de értelmezhetőbb faktorokat[[8]](#footnote-8) kapunk (lásd 1. táblázat). Érdekes eredmény, hogy az elkerülés tételek két faktorba (Faktor1 és Faktor3) rendeződnek, aszerint, hogy fordított tételek-e vagy sem. A szorongás tételek (itt nem voltak fordított megfogalmazásúak) ugyanarra a faktorra (Faktor2) illeszkednek.

**1. Táblázat**

Az ECR-RS Partner skálához tartozó 10 tételének FKA elemzésében a Promax ferde forgatás utáni rendezett faktorsúlymátrix (a 0,2-nél kisebb súlyok helyét üresen hagytuk)

|  |  |  |  |
| --- | --- | --- | --- |
| **Tétel** | **Faktor1** | **Faktor2** | **Faktor3** |
| Elk\_pár1F | 0,908 |  |  |
| Elk\_pár2F | 0,840 |  |  |
| Elk\_pár3F | 0,815 |  |  |
| Elk\_pár4F | 0,771 |  |  |
| Sz\_pár8 |  | 0,888 | -0,204 |
| Sz\_pár9 |  | 0,864 |  |
| Sz\_pár7 |  | 0,817 |  |
| Sz\_pár10 |  | 0,682 |  |
| Elk\_pár6 |  |  | 0,898 |
| Elk\_pár5 |  |  | 0,877 |

*Megjegyzés*: A tételek nevében Elk = Elkerülés, Sz = Szorongás, pár = partner;

Az F betű a tételnév végén azt jelzi, hogy a tétel fordított megfogalmazású.

Ezután – szintén az FKA modul segítségével – reliabilitás-elemzést végeztünk a BDI és a WBI-5 skálára (tételeiket FKA-elemzésre jelölve ki) abból a célból, hogy megtudjuk: kellő megbízhatóságúak-e a jelen vizsgálatban. Ehhez a modul feladatablakában bejelöltük a „Reliabilitási mutatók kiszámítása” opciót, s ezzel kapcsolatban az alábbi eredményeket kaptuk:

* BDI: Cronbach-alfa = 0,816, CI\_0,95 = (0,786; 0,846)
* BDI: McDonald-féle omega = 821, CI\_0,95 = (0,779; 0,863)
* WBI-5: Cronbach-alfa = 0,819, CI\_0,95 = (0,788; 0,850)
* WBI-5: McDonald-féle omega = 0,820, CI\_0,95 = (0,784; 0,857).

Mivel mindkét belső konzisztencia mutató mind a két skála esetén 0,80 feletti, e teszteket a jelen vizsgálatban is kiváló reliabilitásúnak tekinthetjük (DeVellis, 2016).

**3.2. Hierarchikus regresszióelemzés**

Az ECR-RS Partner skála tételeivel végzett FKA elemzésben feltárt faktorok relevanciájának vizsgálatára hierarchikus regresszióelemzést (HierR-t) végeztünk a három faktorral (forgatott főkomponenssel) mint független változóval a Beck-féle Depresszió Skála (BDI) és a WHO Jóllét Skála (WBI-5) összpontszámán, valamint a BFI öt skáláján (Extraverzió, Barátság, Lelkiismeretesség, Érzelmi instabilitás, Nyitottság), a faktorokat egymás után, egyenként léptetve be a regressziós modellbe. Ezen elemzésekben csak a BDI és a WBI-5 esetén kaptunk 0,10 feletti összvarianciát, s az ezekre vonatkozó eredményeket a 2. és a 3. táblázatban foglaltuk össze. Mindkét táblázat arra utal, hogy az ECR-RS Partner skálája esetében az 5. és a 6. elkerülés tétel által kifeszített faktor (Faktor3) nem rendelkezik önálló, plusz szignifikáns prediktív információval a depresszió és a jóllét bejóslásában, ami egyébként a BFI öt skálája esetén is igaz volt.

**2. Táblázat**

Az ECR-RS Partner skála 3 forgatott főkomponensének (Faktor1, Faktor2, Faktor3) hierarchikus regressziója a Beck-féle Depresszió Skála összpontszámára (*F* az *R*2 növekményt tesztelő *F*-statisztika, *f*1 és *f*2 a megfelelő szabadságfokok)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Modell** | ***R*2** | ***R*2 növekmény** | ***F*** | ***f*1** | ***f*2** | ***p*-érték** |
| Faktor1 | 0,1573 |  |  |  |  |  |
| Faktor1, Faktor2 | 0,2190 | 0,0617 | 25,593 | 1 | 324 | < 0,0001 |
| Faktor1, Faktor2, Faktor3 | 0,2237 | 0,0047 | 1,955 | 1 | 323 | 0,163 |

**3. Táblázat**

Az ECR-RS Partner skála 3 forgatott főkomponensének (Faktor1, Faktor2, Faktor3) hierarchikus regressziója a WHO Jóllét Skála összpontszámára (*F* az *R*2 növekményt tesztelő *F*-statisztika, *f*1 és *f*2 a megfelelő szabadságfokok)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Modell** | ***R*2** | ***R*2 növekmény** | ***F*** | ***f*1** | ***f*2** | ***p*-érték** |
| Faktor1 | 0,0728 |  |  |  |  |  |
| Faktor1, Faktor2 | 0,1234 | 0,0506 | 18,684 | 1 | 324 | < 0,0001 |
| Faktor1, Faktor2, Faktor3 | 0,1237 | 0,0003 | 0,125 | 1 | 323 | 0,724 |

**3.3. Feltáró faktorelemzés**

A következő lépésben feltáró faktorelemzéseket végeztünk az ECR-RS 10-10 tételén mind a négy kötődési személy viszonylatában a ROP-R EFA moduljának segítségével. Mivel 7-pontos Likert-skálájú tételek mindegyike *p* < 0,001 szinten szignifikánsan nem normális eloszlású volt (a normalitást a ferdeségi és a csúcsossági együtthatóval tesztelve), EFA módszeréül a főfaktor-elemzést (PAF) választottuk (vö. Vargha, 2019, 105. o.), a megtartott faktorok Promax ferde forgatásával.

A Kaiser-Meyer-Olkin adekvációs mutató (KMO) értéke a négy EFA elemzésben 0,818- 0,847 közé esett. Ez kiváló nagyságszintnek számít (vö. Vargha, 2019, 5.1. táblázat), jelezve, hogy a közösen átfedett rész nagysága alapján mind a négy változóegyüttes alkalmas EFA faktorizációra. Kérdés, hogy hány faktort tartsunk meg, hány faktoros megoldást válasszunk? Ehhez elsőként az FKA-ban kapott sajátértékekre érdemes rápillantani (lásd 4. táblázat).

**4. Táblázat**

Az ECR-RS tételein elvégzett FKA elemzések sajátértékei a négy kötődési személy viszonylatában

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Index** | **Partner** | **Anya** | **Apa** | **Barát** |
| 1 | 4,79 | 4,51 | 4,76 | 4,35 |
| 2 | 1,41 | 2,12 | 1,92 | 1,91 |
| 3 | 1,04 | 0,82 | 1,12 | 0,98 |
| 4 | 0,59 | 0,64 | 0,60 | 0,59 |
| 5 | 0,55 | 0,56 | 0,43 | 0,52 |
| 6 | 0,47 | 0,45 | 0,32 | 0,49 |

**5. Táblázat**

Az ECR-RS tételein PAF módszerrel elvégzett 3-faktoros EFA elemzéseiben a faktorok forgatás előtti megmagyarázott varianciái a négy kötődési személy viszonylatában

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Faktor** | **Partner** | **Anya** | **Apa** | **Barát** |
| Faktor1 | 4,44 | 4,21 | 4,45 | 3,99 |
| Faktor2 | 1,06 | 1,82 | 1,63 | 1,60 |
| Faktor3 | 0,72 | 0,54 | 0,84 | 0,59 |

Az FKA-ban kapott sajátértékek minden viszonylatban azt jelzik, hogy a 4. sajátértéktől kezdve laposodik el a sajátértékek – itt most nem ábrázolt – grafikonja, tehát az első három emelkedik ki a többi közül markánsan, leginkább a Partner és az Apa viszonylatban. A sajátértékek, mint a főkomponensek által a változók varianciájából összesen megmagyarázott varianciaarányok nemcsak a változók közös, hanem egyedi részét, unicitását is tartalmazzák, emiatt az FKA-ban megtartásra érdemes főkomponensek száma felső korlátja az EFA-ban megtartásra érdemes faktorok számának. És mivel ECR-RS skáláit úgy érdemes megszerkeszteni, hogy a skálák száma minden kapcsolati viszonylatban ugyanannyi legyen, a 4. táblázat adatai 2-faktoros struktúrát sejtetnek. Ennek megerősítésére először forgatás nélküli EFA elemzéseket hajtottunk végre 3 faktorral és megnéztük az ezek által megmagyarázott varianciákat (lásd 5. táblázat), melyek a ROP-R eredménylistán a *Rotálatlan faktorsúlyok mátrixa* című táblázat alatt láthatók. Az 5. táblázat megerősíti, hogy a 10 tétel minden viszonylatban kétfaktoros struktúrára illeszkedik a legjobban, ugyanis mindenütt pontosan két faktor megmagyarázott varianciája haladja meg az 1 változónyi varianciát, 1-et.

A faktorszámon túl az EFA elemzésekből azt is szeretnénk megtudni, hogy melyek a két faktorra legjobban illeszkedő tételek a négy viszonylatban. E célból a ROP-R EFA moduljának segítségével EFA elemzéseket végeztünk, PAF módszerét választva és minden esetben 2 faktort forgatva Promax-szal. A négy kötődési személy viszonylatában az eredmény meglehetősen egységes volt. Az első faktort (Faktor1) minden esetben az elkerülés tételek, a másodikat (Faktor2) pedig a szorongás tételek alkották. Ennek szemléltetésére a forgatás utáni rendezett faktorsúlymátrixot a 6. táblázatban mutatjuk be a Barát skálához tartozó 10 tétel EFA elemzésében.

**6. Táblázat**

Az ECR-RS Barát skálához tartozó 10 tétel EFA elemzésében a Promax ferde forgatás utáni rendezett faktorsúlymátrix (a 0,2-nél kisebb súlyok helyét üresen hagytuk)

|  |  |  |
| --- | --- | --- |
| **Tétel** | **Faktor1** | **Faktor2** |
| Elk\_Ba2F | 1,020 | 0,246 |
| Elk\_Ba3F | 0,975 | 0,241 |
| Elk\_Ba1F | 0,629 |  |
| Elk\_Ba4F | 0,589 |  |
| Elk\_Ba5 | -0,559 |  |
| Elk\_Ba6 | -0,523 |  |
| Sz\_Ba7 |  | 0,857 |
| Sz\_Ba8 |  | 0,796 |
| Sz\_Ba9 |  | 0,793 |
| Sz\_Ba10 | -0,325 | 0,431 |

*Megjegyzés*: A tételek nevében Elk = Elkerülés, Sz = Szorongás, Ba = Barát;

Az F betű a tételnév végén azt jelzi, hogy a tétel fordított megfogalmazású

A 6. táblázatból láthatjuk, hogy az elkerülés tételek közül Faktor1-re az első négy tétel illeszkedik a legjobban, s ez ugyanígy megfigyelhető volt az Apa és a Partner viszonylatban is. Az Anya viszonylatban annyi eltérés volt, hogy a fordított megfogalmazású 5. és 6. tétel (ezért ellentétes ezek faktorsúlyának az előjele a másik négy tételével) a többi viszonylatnál sokkal erősebben (0,70 feletti súlyokkal) illeszkedett Faktor1-re. A többi viszonylatban még olyan is előfordult (konkrétan a Partner skálánál), hogy az 5. és 6. elkerülés tétel súlya 0,50 alá esett (-0,475 és -0,371), de azért skálaképző nagyságú maradt. Mindezek alapján, azt is figyelembe véve, hogy a kisszámú fordított tételek gyakran gyengébben illeszkednek ugyanarra a faktorra, mint a nem fordítottak, az Elkerülés skála esetén érdemes meghagynunk mind a 6 tételt.

A 6. táblázatból azt is kiolvashatjuk, hogy a Szorongás skálára a 7-9. tétel illeszkedik a legjobban, s ez ugyanígy megfigyelhető volt minden más kötődési viszonylatban is, azzal együtt, hogy az utolsó, 10. tétel minden esetben kilógott a skálából, három esetben 0,50 alatti (0,272; 0,405; 0,431) súlyokkal. Egyedül a Partner viszonylatban haladta meg a faktorsúly a 0,50-et (0,594), de itt is kisebb volt, mint a másik három tételé. Mindezek alapján érdemes a Szorongás skála esetén a 7-9. tételekre szorítkozni, ahogy ezt Fraley és munkatársai (2011) is megállapították.

**3.4. Megerősítő faktorelemzés**

A következő lépésben CFA segítségével ellenőriztük, hogy az ECR-RS Elkerülés és Szorongás skálája a fenti faktormodellre megfelelően illeszkedik-e magyar mintánkban. A normalitás sérülése miatt az illesztéshez ML robusztus változatát, az MLMV becslési módszert választottuk, mely több szerző által jó alternatívája ML-nek (vö. Gao, Shi és Maydeu-Olivares, 2020; Zábó, Oláh és Vargha, 2022). Ehhez ROP-R CFA moduljában mind a négy kapcsolati viszonylat esetében az 1-6. tételek blokkindexét meghagytuk 1-nek, a 7-9. tételekét pedig 2-re állítottuk be. Az első futás során minden viszonylatban az Elkerülés skála 5. és 6. (5. Nem szívesen nyílok meg neki, 6. Inkább nem mutatom ki neki, hogy mit érzek mélyen a szívemben), illetve a 2. és 3. (2. Általában megbeszélem a gondjaimat, problémáimat vele, 3. Átbeszéljük a dolgokat vele) tétele esetében nőtt a Modifikációs index olyan magasra (35 fölé), mely azt jelezte, hogy ezeket a faktoron belüli kovarianciákat érdemes a modellbe beépíteni. A kovarianciák beépítésének és a robusztus becslés alkalmazásának sikerességét az jelzi, hogy ekkor a faktormodell tesztelése egyetlen esetben sem volt szignifikáns *p* < 0,001 szinten (lásd 7. táblázat, MLMV+ oszlop), míg a többi esetben egyetlen kivételtől eltekintve (lásd 7. táblázat, Anya viszonylat, ML+ oszlop) mindig.

**7. Táblázat**

Az ECR-RS faktormodelljének tesztelése χ2-próbával a normalitást feltételező ML és a robusztus MLMV módszerrel a kovarianciák beépítése előtt (ML és MLMV oszlop) és után (ML+ és MLMV+ oszlop) a négy kötődési személy viszonylatában

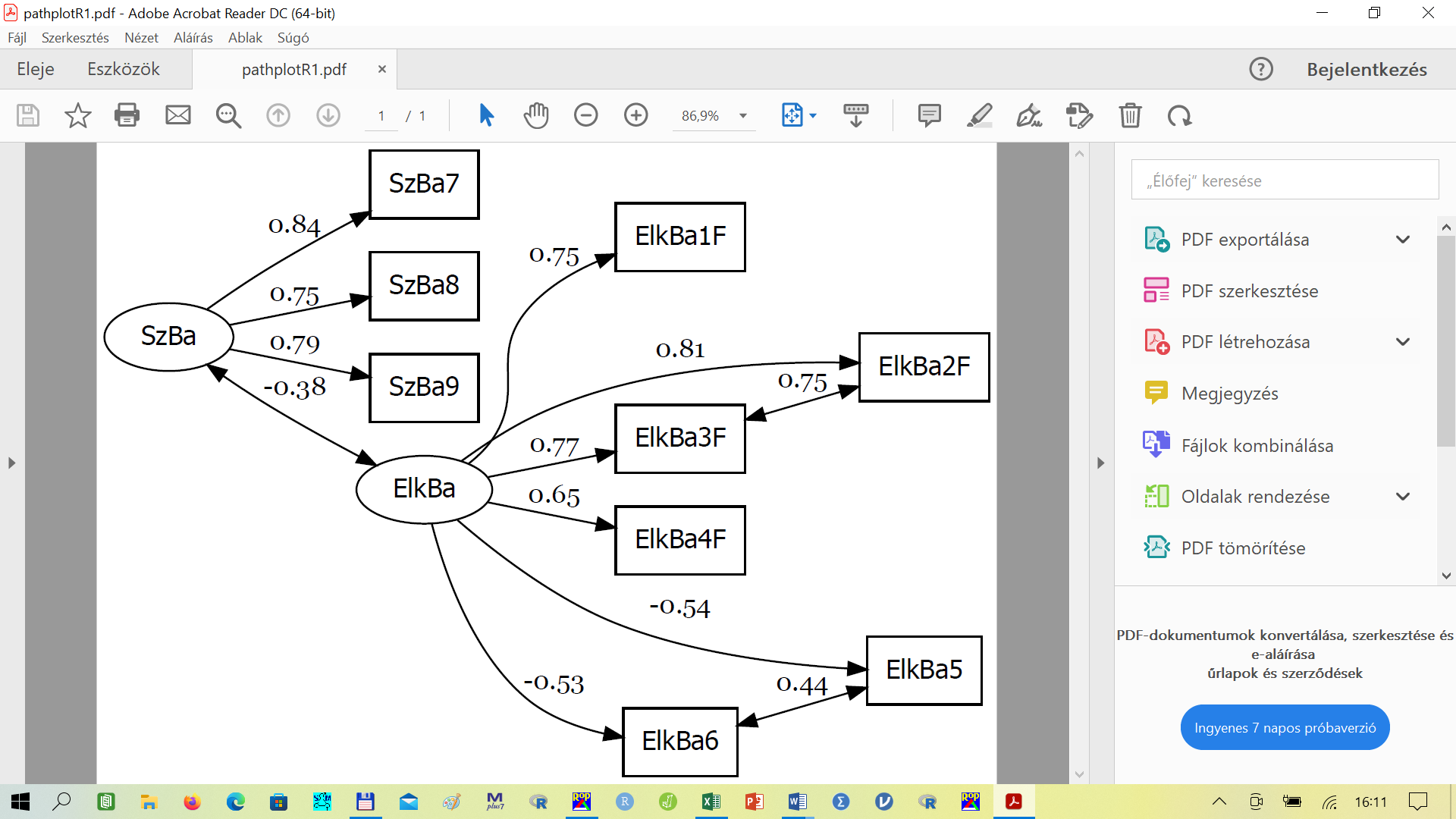
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Mutató** | **ML** | **MLMV** | **ML+** | **MLMV+** |
| Partner | χ2 | 166,12 | 103,4 | 62,19 | 41,92 |
|  | *f* | 26 | 26 | 24 | 24 |
|  | *p*-érték | < 0,001 | < 0,001 | < 0,001 | 0,013 |
| Anya | χ2 | 174,69 | 117,61 | 39,29 | 29,69 |
|  | *f* | 26 | 26 | 24 | 24 |
|  | *p*-érték | < 0,001 | < 0,001 | 0,025 | 0,195 |
| Apa | χ2 | 251,98 | 184,54 | 58,64 | 46,74 |
|  | *f* | 26 | 26 | 24 | 24 |
|  | *p*-érték | < 0,001 | < 0,001 | < 0,001 | 0,004 |
| Barát | χ2 | 169,81 | 104,57 | 50,78 | 35,57 |
|  | *f* | 26 | 26 | 24 | 24 |
|  | *p*-érték | < 0,001 | < 0,001 | 0,001 | 0,060 |

**8. Táblázat**

Az ECR-RS kovarianciákkal javított kétdimenziós faktormodelljének illeszkedési mutatói a négy kötődési viszonylatban, robusztus MLMV becslési módszert alkalmazva

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Mutató** | **Partner** | **Anya** | **Apa** | **Barát** |
| RMSEA | 0,048 | 0,027 | 0,055 | 0,039 |
| CI\_0,95 (RMSEA) | (0,022; 0,071) | (0,000; 0,055) | (0,031; 0,079) | (0,000; 0,064) |
| pClose | 0,532 | 0,899 | 0,326 | 0,738 |
| CFI | 0,975 | 0,992 | 0,980 | 0,985 |
| TLI | 0,962 | 0,989 | 0,971 | 0,978 |
| SRMR | 0,041 | 0,040 | 0,035 | 0,043 |

Az ECR-RS kovarianciákkal javított kétdimenziós faktormodelljének illeszkedési mutatóit a négy kötődési személy viszonylatában a 8. táblázat mutatja be. Ebből azt látjuk, hogy az illeszkedés minden mutató és minden viszonylat esetén jó vagy kiváló (vö. Hu és Bentler, 1999; Browne és Cudeck, 1993; Vargha, 2019, 6.3. táblázat). Ennek a modellnek a path diagramját szemléltetésképpen a Barát kapcsolati viszonylatban az 1. ábra mutatja be. Ezt az ábrát a ROP-R pdf fájlban készíti el automatikusan a CFA modulban (pathplotR1.pdf néven), melyet a <https://pdf2jpg.net/> weblapon lehet ingyenesen jpg formátumúra konvertálni.

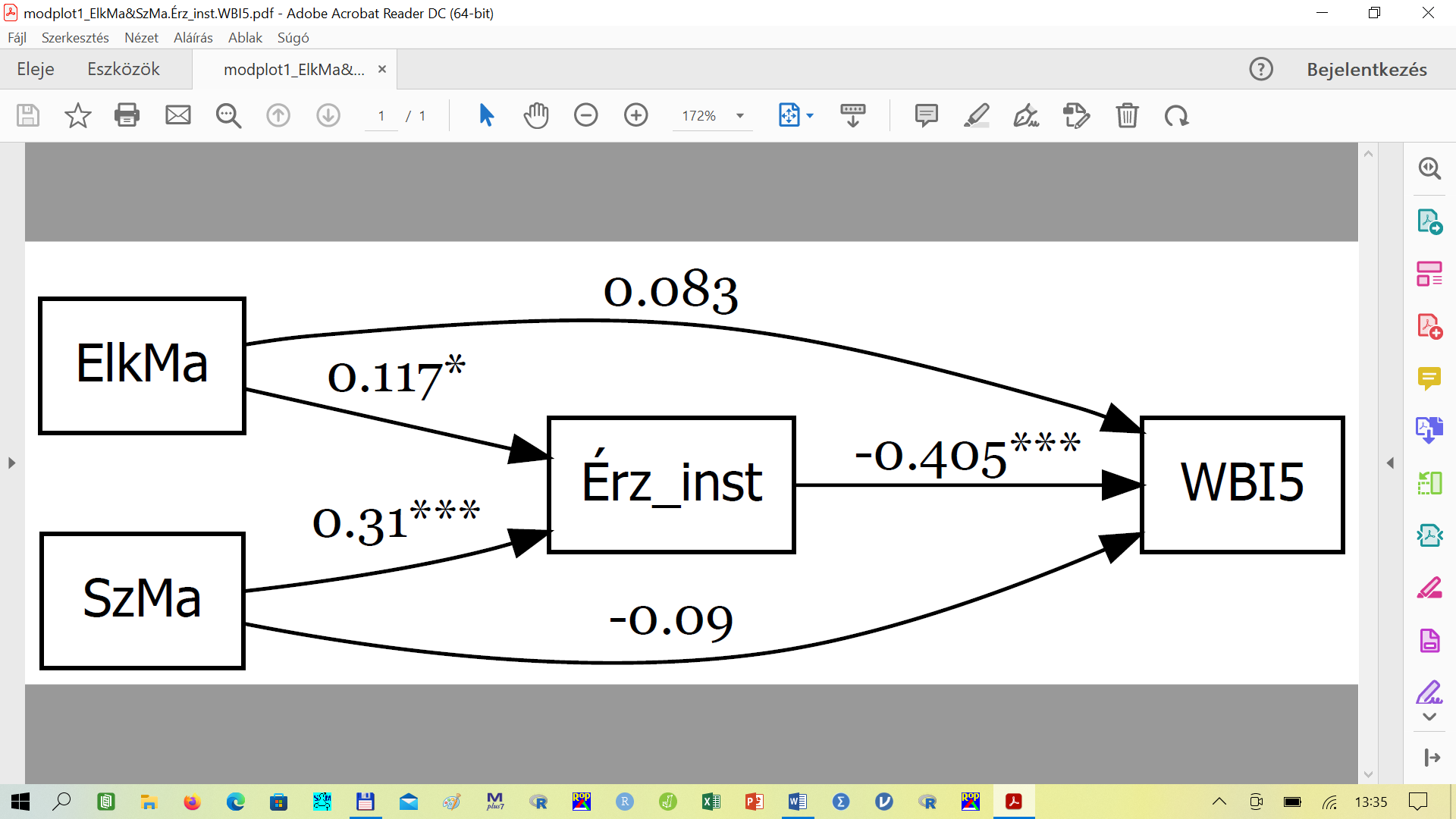


**1. Ábra:** Az ECR-RS kovarianciákkal javított kétdimenziós modelljének path diagramja standardizált regressziós becslésekkel a Barát kötődési viszonylatban (Sz = Szorongás, Elk = Elkerülés, Ba = Barát)

A CFA modullal végzett elemzések alapján összegzésképpen megállapíthatjuk, hogy az ECR-RS jó strukturális validitású. Mindezek után érdemes a CFA modulban létre is hoznunk a skálákat, mely a „Faktorindexekkel definiált skálák létrehozása” opció bejelölésével végezhető el a legegyszerűbben. Ennek hatására ROP-R a skálákat az alkotó tételek átlagaként hozza létre. A fordított tételekkel se kell bajlódnunk, mert azokat a ROP-R a faktorsúlyok előjele alapján azonosítja és átfordítja.

**3.5. Mediációs elemzés**

A strukturális validitáson túl egy tesztadaptáció során a tartalmi validitás különböző formáit is illik igazolni. Ehhez érdemes megvizsgálni mintánkban a létrehozott ECR-RS skálák és a többi teszt kapcsolatát. Ez különböző regressziós elemzésekkel is elvégezhető a ROP-R HierR modulja segítségével, amelyek közül most szemléltetésképpen egy mediációs elemzést mutatunk be. Számos kutatás hívta fel a figyelmet arra, hogy a kötődés szoros kapcsolatban van a mentális egészséggel. A problémás kötődésű személyek, akik magas elkerülési motivációval és erős szorongással rendelkeznek, általában rosszabb mentális egészségűek (lásd pl. Adams, Wrath és Meng, 2018), akiknek a szubjektív jólléte is szükségképpen alacsonyabb. Kérdés, hogy a kötődés kapcsolata a szubjektív jólléttel közvetlen hatás eredménye, vagy esetleg szerepet játszik itt a személyiség valamelyik komponense is. Ennek ellenőrzésére olyan mediációs elemzéseket végeztünk, amelyekben a független változó az ECR-RS Elkerülés és Szorongás skálája volt (külön-külön minden kötődési viszonylatban), a függő változó a WHO Jóllét Skála, a mediátor változók pedig a BFI Big Five teszt skálái.



**2. Ábra:** Az ECR-RS Anyai elkerülés (ElkMa) és Anyai szorongás (SzMa) skálájának kapcsolata a szubjektív jólléttel (WBI5) az érzelmi instabilitás (Érz\_inst) mediációs hatásának figyelembevételével (path diagram)

A legérdekesebb eredményt akkor kaptuk, amikor az anyai kötődés hatását vizsgáltuk a szubjektív jóllétre az érzelmi instabilitás mediációs hatásának figyelembevételével. A 2. ábrán látható diagram arról tájékoztat, hogy az anyai kötődés viszonylatában sem az elkerülés, sem a szorongás nem hat közvetlenül a jóllétre, csupán az érzelmi instabilitás személyiségvonáson keresztül. Eszerint ha egy személynél magas szintű ugyan az anyai elkerülés és szorongás, de nem jellemző rá túlzott érzelmi instabilitás, akkor szubjektív jólléte akár magas szintű is lehet.

Megjegyezzük, hogy a HierR modulban ROP-R akkor kérdez rá, hogy végrehajtson-e mediációs elemzést, ha a független változókat pontosan két blokkba soroljuk (pl. 1 és 2 index-szel), a kisebbik (pl. 1-es) blokkindexűek száma nem haladja meg a 2-t (vagyis 1 vagy 2), és a függő változók száma 10-nél kisebb. Ha a válaszunk igen, akkor a mediációs elemzés a kisebbik indexű blokk hatását fogja vizsgálni minden egyes függő változóra az összes nagyobbik blokkindexű változóval mint mediátor változóval külön-külön elemzésekben.

**3.6. Polinomiális regresszióelemzés**

Felnőtteknél a kötődésnek alapvetően négy különböző típusa létezik, melyeket az elkerülés és a szorongás dimenzió alacsony és magas szintjének négy kombinációja definiál (Jantek és Vargha, 2016). Ezt mintánkban az ECR-RS Elkerülés és Szorongás skálájával lehet kötődési típusonként ellenőrizni. Ha az input változók többdimenziós normális eloszlásúak, az eloszlásnak csak egyetlen csúcsa lesz, így nem lesz esély négy egymástól jól elkülönülő típus feltárására (vö. Vargha, 2020, 10. o.). Mivel normális eloszlású változók között csak lineáris kapcsolatok fordulhatnak elő, a nemlineáris összefüggések kizárják a többdimenziós normalitást és esélyt adnak a mintában több centrum és ezáltal több típus feltárására. A nemlineáris kapcsolatok feltárására polinomiális regresszióelemzést végeztünk minden kötődési viszonylatban az Elkerülés és a Szorongás skála között, mind a két változót mind független, mind függő változó szerepben alkalmazva. A nemlineáris kapcsolatot az mutatja, ha a független változónak a nemlineáris hatásokat képviselő magasabb (1-nél nagyobb) hatványai szignifikáns predikciós hatásúak a függő változó bejóslásában.

Elemzéseink azt mutatták, hogy nemlineáris hatás csak a Partner viszonylatban fordul elő az Elkerülés és a Szorongás skála között, itt viszont mindkét irányban. Az ezzel kapcsolatos legfontosabb eredményeket a 9. és a 10. táblázat tartalmazza.

**9. Táblázat**

Az Elkerülés skála polinomiális regressziója a Szorongás skálára a Partner viszonylatban

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Modell** | ***R*2** | ***R*2 növekmény** | ***F*** | ***f*1** | ***f*2** | ***p*-érték** |
| Lineáris | 0,063 |  |  |  |  |  |
| Kvadratikus | 0,069 | 0,007 | 2,29 | 1 | 326 | 0,131 |
| Harmadfokú | 0,127 | 0,058 | 21,493 | 1 | 325 | < 0,0001 |
| Negyedfokú | 0,140 | 0,013 | 4,889 | 1 | 324 | 0,028 |
| Ötödfokú | 0,146 | 0,006 | 2,341 | 1 | 323 | 0,127 |

**10. Táblázat**

A Szorongás skála polinomiális regressziója az Elkerülés skálára a Partner viszonylatban

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Modell** | ***R*2** | ***R*2 növekmény** | ***F*** | ***f*1** | ***f*2** | ***p*-érték** |
| Lineáris | 0,063 |  |  |  |  |  |
| Kvadratikus | 0,069 | 0,007 | 2,379 | 1 | 326 | 0,124 |
| Harmadfokú | 0,079 | 0,010 | 3,531 | 1 | 325 | 0,061 |
| Negyedfokú | 0,100 | 0,021 | 7,44 | 1 | 324 | 0,007 |
| Ötödfokú | 0,100 | 0,000 | 0,144 | 1 | 323 | 0,704 |

Eredményként azt látjuk, hogy a Szorongás skála bejóslásában az Elkerülés skála harmadfokú komponensének van a legmarkánsabb – nemlineáris – hatása (*p* < 0,0001), az Elkerülés skála bejóslásában pedig a Szorongás skála negyedfokú komponensének (*p* = 0,007). Látható, hogy a szignifikáns nemlineáris hatások számottevően megemelik *R***2**-et (0,063-ról 0,140-re, illetve 0,063-ról 0,100-ra). Mindezen eredmények adnak esélyt arra, hogy klaszterelemzéssel feltárhatók legyenek a kötődés ismert típusai. Ennek sikerességét mutatják Jantek és Vargha (2016) elemzései, továbbá Vargha és Bánsági (2022) ROP-R segítségével elvégzett klaszterelemzései.

**Megbeszélés**

Cikkünkben a Windows rendszerben futtatható, ingyenes, kétnyelvű (magyar és angol) és egyszerűen használható ROP-R többváltozós statisztikai programcsomagot mutattuk be, mely jól használható pszichológiai kutatások statisztikai elemzéséhez. Felhasználó-barát menüjében ROP-R a többváltozós statisztika három fontos témakörében (regresszióelemzés, főkomponens- és faktoranalízis, illetve klaszteranalízis) kínál teljes körű elemzéseket. A JASP és a jamovi szoftverhez hasonlóan ROP-R a statisztikai menüjében kiválasztott és beállított elemzésekhez R-scripteket ír, azokat automatikusan lefuttatja, majd a kapott eredményeket táblázatos formában ROP-R nézőkéjében megjeleníti, illetve az eredményekhez tartozó fontosabb grafikonokat jpg vagy pdf fájlokban elmenti. ROP-R moduljai számos olyan fontos statisztikai elemzés (pl. mediációs és polinomiális regresszióelemzés, *k*-medoid és *k*-medián nemhierarchikus klaszterelemzés vagy modell-alapú klaszteranalízis) végrehajtását teszik lehetővé, amelyek más felhasználó-barát szoftverekben jelenleg nem elérhetők. ROP-R standard installálása az alábbi lépésekkel hajtható végre.

* Telepítse ROP-R-t a <http://www.bansagi.hu/r/index.html> weboldalról. Sikeres installálás után a ROP-R.exe indítóprogram a c:\\_vargha\ropstat mappában található.
* Telepítse az R szoftver Windows rendszerekre kidolgozott R-4.1.3 verzióját a következő weboldalról: https://www.filehorse.com/download-r-for-windows/old-versions/ (ha még nem lett korábban a gépére telepítve). Az R újabb (4.2.0 és efölötti) verzióival a ROP-R nem fut normálisan.
* Indítsa el a ROP-R.exe programot és állítsa be az Rcmd.exe program elérési útját a ROP-R Beállítások/R-path menüpontjában. Standard telepítés során ez az elérési út: c:\Program Files\R\R-4.1.3\bin\i386\Rcmd.exe, de a c:\Program Files\R\R-4.1.3\bin\x64\Rcmd.exe szintén használható. Utóbbi 64 bites eljárások esetén gyorsabb futásokat eredményez, mint az előbbi.
* Indítsa el az RGui.exe programot (nincs parancsikonja R-nek az asztalon vagy a tálcán, akkor RGui.exe ugyanabban a mappában keresendő, melyben az Rcmd.exe programot kijelöltük).
* Ha a gépen korábban már használták RGui-t és installáltak benne R-package-et, akkor simán másolja be a cikk 1. fejezetének 3.b alpontjában felsorolt package-installáló R-utasításokat és Enter-rel futtassa le őket.
* Ha a gépén korábban nem volt R telepítve vagy volt, de RGui-ben egyetlen package-et se installáltak, akkor a fentebbi utasításban említett utasításlistából először csak az elsőt másolja be és futtassa RGui-ben, fogadja el az ott felkínált lehetőségeket, majd sikeres futtatás után ugyanezt tegye a többivel, de ne egyenként, hanem egy csomagban.

A fenti lépéseket a ROP-R használatához csupán egyszer kell megtennie. Sikeres végrehajtásuk után élvezheti a ROP-R cikkünk 2. fejezetében részletezett tíz moduljának eleganciáját és hatékonyságát, amit a 3. fejezetben egy kötődéskutatás adatain végzett komplex statisztikai elemzéssorral szemléltettünk.

**Köszönetnyilvánítás**

A jelen cikk elkészítését a Károli Gáspár Református Egyetem Bölcsészet- és Társadalomtudományi Karának *Pszichológiai kutatások módszertani platformja* című kutatói pályázata támogatta (témaszám: 20754B800/2022). A szerzők ezúttal mondanak köszönetet a kézirat lektorainak és a folyóirat felelős szerkesztőjének számtalan értékes megjegyzéséért és javaslatáért, amelyek nagyban segítették a cikk végső változatának megformálását.

**Irodalom**

Adams, G. C., Wrath, A. J., & Meng, X. (2018). The relationship between adult attachment and mental health care utilization: A systematic review. *The Canadian Journal of Psychiatry*, *63*(10), 651-660.

Bech, P. (2012). *The Bech, Hamilton and Zung scales for mood disorders: screening and listening: a twenty years update with reference to DSM-IV and ICD-10*. Springer Science & Business Media.

Bech, P. (1996). *The Bech, Hamilton and Zung scales for mood disorders: screening and listening. 2nd ed.* Berlin: Springer.

Beck, A. T., & Beck, R. W. (1972). Screening depressed patients in family practice: A rapid technic. *Postgraduate medicine*, *52*(6), 81-85.

Bergman, L. R., Magnusson, D., & El-Khouri, B. M. (2003). *Studying individual development in an interindividual context. A Person-oriented approach.* Lawrence-Erlbaum. Associates.Mahwa, New Jersey, London.

Bergman, L. R., Vargha, A., & Kövi, Z. (2017). Revitalizing the typological approach: Some methods for finding types. *Journal for Person-Oriented Research, 3*(1), 49−62. DOI: 10.17505/jpor.2017.04, <https://www.person-research.org/journal/files/3_1/filer/4.pdf>

Bernaards, C. A., & Jennrich, I. R. (2005). Gradient Projection Algorithms and Software for Arbitrary Rotation Criteria in Factor Analysis. *Educational and Psychological Measurement, 65*(2), 676–696.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J.S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Sage.

Cardot, H. (2022). *Gmedian: Geometric Median, k-Medians Clustering and Robust Median PCA.* R package version 1.2.7. https://CRAN.R-project.org/package=Gmedian

DeVellis, R. F. (2016). *Scale development: Theory and applications (4th ed.).* Sage Publications.

Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American statistical Association*, *97*(458), 611-631.

Fraley, R. C., Heffernan, M. E., Vicary, A. M., & Brumbaugh, C. C. (2011). The experiences in close relationships—Relationship Structures Questionnaire: A method for assessing attachment orientations across relationships. *Psychological assessment*, *23*(3), 615.

Gao, C., Shi, D., & Maydeu-Olivares, A. (2020). Estimating the maximum likelihood root mean square error of approximation (RMSEA) with non-normal data: A Monte-Carlo study. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(2), 192-201.

Gergely, B., & Vargha, A. (2021). How to Use Model-Based Cluster Analysis Efficiently in Person-Oriented Research. *Journal for Person-Oriented Research*, *7*(1), 22. DOI: 10.17505/ jpor.2021.23449 <https://journals.lub.lu.se/jpor/article/view/23449/20820>

Harrington, D. (2009). Confirmatory factor analysis. Oxford: Oxford university press.

Hebbali, A. (2020). *olsrr: Tools for Building OLS Regression Models.* R package version 0.5.3. <https://CRAN.R-project.org/package=olsrr>

Henry, L., & Wickham, H. (2022). rlang: Functions for Base Types and Core R and 'Tidyverse' Features. R package version 1.0.5. https://CRAN.R-project.org/package=rlang

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1-55. <https://doi.org/10.1080/10705519909540118>

Jantek, G., & Vargha, A. (2016). A felnőtt kötődés korszerű mérési lehetősége: A közvetlen kapcsolatok élményei—kapcsolati struktúrák (ECR-RS) kötődési kérdőív magyar adaptációja párkapcsolatban élő felnőtt személyeknél. *Magyar Pszichológiai Szemle*, *71*(3), 447-470. DOI: <http://dx.doi.org/10.1556/0016.2016.71.3.3>

JASP Team (2022). JASP (Version 0.16.2) [Computer software]. <https://jasp-stats.org/>

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). The Big Five Inventory - Versions 4a and 54. Berkeley: University of California, Institute of Personality and Social Research.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins & L. A. Pervin, *Handbook of personality: Theory and research*. New York: Guilford Press, 114-158

de Jong, P. F. (1999). Hierarchikus regresszió analysis in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(2), 198-211.

Kassambara, A., & Mundt, F. (2020). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses.* R package version 1.0.7. https://CRAN.R-project.org/package=factoextra

Kaufman, L. & Rousseeuw, P. J. (2009). *Finding Groups in Data: an Introduction to Cluster Analysis.* New York: John Wiley & Sons.

Kelley, K. (2007). *Methods for the Behavioral, Educational, and Social Sciences (MBESS)* [Computer software and manual]. <https://CRAN.R-project.org/package=MBESS>

Kopp, M., Skrabski, Á., & Czakó, L. (1990). Összehasonlító mentálhigiénés vizsgálatokhoz ajánlott módszertan. *Végeken, 1*(2), 4-24.

Leys, C., Klein, O., Dominicy, Y., & Ley, C. (2018). Detecting multivariate outliers: Use a robust variant of Mahalanobis distance. *Journal of Experimental Social Psychology, 74*, 150-156. https://doi.org/10.1016/j.jesp.2017.09.011

Lishinski, A. (2021). *lavaanPlot: Path Diagrams for 'Lavaan' Models via 'DiagrammeR'.* R package version 0.6.2. <https://CRAN.R-project.org/package=lavaanPlot>

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2022). *cluster: Cluster Analysis Basics and Extensions.* R package version 2.1.4. <https://CRAN.R-project.org/package=cluster>

Marzjarani, M. (2015). Sample size and outliers, leverage, and influential points, and Cooks distance formula. *International Journal of Arts and Commerce, 4*(2), 83-86.

Mouselimis, L. (2022). *ClusterR: Gaussian Mixture Models, K-Means, Mini-Batch-Kmeans, K-Medoids and Affinity Propagation Clustering.* R package version 1.2.6, <https://CRAN.R-project.org/package=ClusterR>

Osborne, J. W. (2014). *Best Practices in Exploratory Factor Analysis.* Scotts Valley, CA: CreateSpace Independent Publishing.

Peng, R. D. (2016). *R programming for data science*. Victoria, BC, Canada: Leanpub, 86-181. <https://bookdown.org/rdpeng/rprogdatascience/>

R Core Team (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

Revelle, W. (2022). *psych: Procedures for Personality and Psychological Research*. Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psych Version = 2.2.5.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software,* *48*(2), 1-36. https://doi.org/10.18637/jss.v048.i02

Rosseel, Y. és mtársai (2018). lavaan: Latent Variable Analysis. [R package]. Letöltés: https://cran.r-project.org/package=lavaan

Roux, M. (2018). A comparative study of divisive and agglomerative hierarchical clustering algorithms. *Journal of Classification*, *35*(2), 345-366.

Rózsa, S., Tárnok, Z., & Nagy, P. (2020). *A gyermekpszichiátriában alkalmazott kérdőívek, interjúk és tünetbecslő skálák.* Budapest: EFOP-2.2.0-16.2016.00002 Gyermek és ifjúságpszichiátriai, addiktológiai és mentálhigiénés ellátórendszer infrastrukturális feltételeinek fejlesztése projekt.

Şahin, M., & Aybek, E. (2019). Jamovi: an easy to use statistical software for the social scientists. *International Journal of Assessment Tools in Education*, *6*(4), 670-692.

Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal.* Vol. 8. No. 1. pp. 289-317.

T. Kárász, J., Nagybányai Nagy, O., Széll, K., & Takács, S. (2022). Cronbach-alfa: vele vagy nélküle? *Magyar Pszichológiai Szemle*, *77*(1), 81-98. Letöltés: https://akjournals.com/view/journals/0016/77/1/article-p81.xml

Tabachnick, B. G., & Fidell, L. S. (2013). Using multivariate statistics (6th ed.). Boston: Pearson.

The jamovi project (2021). jamovi (Version 1.6) [Computer Software]. Letöltés: <https://www.jamovi.org>

Vargha, A. (2016). A ROPstat statisztikai programcsomag. *Statisztikai Szemle*, *94*(11-12)*,* 1165–1192. DOI: 10.20311/stat2016.11-12.hu1165, http://www.ksh.hu/statszemle\_archive/2016/2016\_11-12/2016\_11-12\_1165.pdf

Vargha, A. (2019). *Többváltozós statisztika dióhéjban: változó-orientált módszerek.* Pólya Kiadó, Budapest. ISBN 978-615-81228-1-8

Vargha, A. (2020). *Normális vagy? És ha nem? Statisztikai módszerek nem normális eloszlású változókkal pszichológiai kutatásokban.* Pólya Kiadó, Budapest. ISBN 978-615-81228-3-2

Vargha, A. (2021). *Személy-orientált többváltozós statisztika: klasszifikációs módszerek.* Pólya Kiadó, Budapest. ISBN 978-615-81228-4-9

Vargha, A., & Bánsági, P. (2022). ROP-R: a free multivariate statistical software that runs R packages in a ROPstat framework. *Hungarian Statistical Review*, *5* (in print)

Vargha, A., & Bergman, L. R. (2019). MORI coefficients as indicators of a ‘real’ cluster structure. *Hungarian Statistical Review*, *2*(1), 3–23. DOI: 10.35618/hsr2019.01.en003, http://real.mtak.hu/95789/

Vargha, A., Bergman, L. R., & Takács, S. (2016). Performing cluster analysis within a person-oriented context: Some methods for evaluating the quality of cluster solutions. *Journal for person-oriented research*, *2*(1-2), 78–86. DOI: 10.17505/jpor.2016.08. https://www.person-research.org/journal/files/2\_1/filer/22.pdf

Vargha, A., Torma, B., & Bergman, L. R. (2015). ROPstat: a general statistical package useful for conducting person-oriented analyses. *Journal for Person-Oriented Research*, *1*(1-2), 87–98. DOI: 10.17505/jpor.2015.09 https://www.person-research.org/journal/files/1\_1/filer/44.pdf

Venables, W. N., & Ripley, B. D. (2002). Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis.* Springer Verlag, New York.

Zábó, V., Oláh, A., & Vargha, A. (2022). A new complex mental health test in positive psychological framework. *Frontiers in Psychology*, Vol. 13. No. 775622. doi: 10.3389/fpsyg.2022.775622 https://www.frontiersin.org/articles/10.3389/fpsyg.2022.775622/full

**Psychological statistical analyses with the ROP-R software**

**Abstract**

The present paper introduces ROP-R, a new and free ROPstat based multivariate statistical program package. ROP-R provides a user-friendly menu system for running several statistical procedures (several regressions, dimension reduction analyses, clustering procedures) by means of ROP-R created R scripts, just as JASP or jamovi, but in several modules offering new opportunities (e.g., model-based cluster analysis). The usefulness and elegance of ROP-R is illustrated in chapter 3 of the present paper by a complex series of statistical analyses made with a data sample from a study in the psychological topic of attachment.

**Keywords**: ROP-R, ROPstat, R-scripts

1. Elkészült kézirat az Alkalmazott Pszichológia folyóiratban való megjelentetéshez [↑](#footnote-ref-1)
2. Kivéve az üzleti célú vállalkozásokat [↑](#footnote-ref-2)
3. https://cran.r-project.org/web/packages/available\_packages\_by\_name.html [↑](#footnote-ref-3)
4. Vö. <https://askubuntu.com/questions/54296/difference-between-the-i386-download-and-the-amd64> [↑](#footnote-ref-4)
5. Excel fájlok beolvasásakor az adatfájlt tartalmazó munkalapot kell aktív lapnak beállítani. [↑](#footnote-ref-5)
6. A dummy változó értéke egy személynél 1, ha a személyt az adott érték jellemzi, egyébként 0. [↑](#footnote-ref-6)
7. A forgatott főkomponenseket szokták faktoroknak is nevezni. [↑](#footnote-ref-7)
8. A forgatott főkomponenseket az egyszerűség kedvéért itt most faktoroknak nevezzük. [↑](#footnote-ref-8)