**András Vargha – Péter Bánsági**

# ROP-R: a free multivariate statistical software that runs R packages in a ROPstat framework[*]

András Vargha, Professor Emeritus, Károli Gáspár University of the Reformed Church, Professor Emeritus, Eötvös Loránd University, Budapest, Hungary
Email: vargha.andras@kre.hu
Péter Bánsági, Budapest, Hungary
Email: peter@bansagi.hu

This 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 the ROP-R are illustrated in a separate chapter by a complex series of statistical analyses made with a data sample from a study on the psychological topic of attachment.

Keywords: ROP-R, ROPstat, R-scripts

ROPstat is one of the few Hungarian developments in wide-scope statistical software that has attained some international reputation *(Vargha et al., 2015).* It specialises in robust techniques, ordinal analyses, and pattern- and person-oriented methods, many of which are not available in other common statistical software. Despite its merits, ROPstat users may feel bothered by the lack of basic multivariate methods, such as nonlinear regression or dimension reduction methods (e.g. principal component analysis and factor analysis).

Paying software (SPSS, SAS, etc.) dominated the statistical software market several decades ago. However, approximately twenty years ago, the appearance of R, an open-source free software, started to change the minds of software developers and users. Almost all statistical methods can be performed using one or more R packages. The only difficulty with R is that any analysis in R can only

---

[*] The preparation of the present paper was supported by a research grant obtained from the Faculty of Humanities and Social Sciences, Károli Gáspár University of the Reformed Church, titled *Methodological platform of psychological studies* (Grant No. 2022/20754B800).

be performed with a precisely prepared set of R commands, called scripts, that satisfy severe syntax rules. For example, lower- and upper-case letters are seldom interchangeable in R commands.

To overcome this difficulty, two user-friendly statistical software programs, JASP *(JASP Team, 2022)* and jamovi *(The jamovi project, 2021; Şahin–Aybek, 2019)* have been developed recently. Both projects were founded to develop a free and open statistical platform that is intuitive to use and can provide the latest developments in statistical methodology. Their common philosophy is that scientific software should be community-driven, wherein anyone can develop and publish analyses and make them freely available to a wide audience.[1] Both software programs are based on R but provide a user-friendly menu system for large-scale statistical analyses without the need for knowledge of the sophisticated syntax rules of R.

Following this philosophy, the developers of ROPstat – the authors of the present paper – decided to develop ROP-R, a multivariate extension of ROPstat with the following characteristics: 1) ROP-R is a ROPstat-based free multivariate statistical program package; 2) ROP-R provides a user-friendly menu system for running several statistical procedures by means of the software-created R scripts, just as JASP or jamovi. The first, 1.0 version of the ROP-R was launched in May 2022, and can be downloaded from http://www.bansagi.hu/r/index.html. The main features of ROP-R are as follows:

1. It is a free, bilingual (English and Hungarian) multivariate statistical software that provides analyses in its present version in three types of statistical modules:
   - Regression,
   - Dimension reduction,
   - Clustering.
2. In each module, the user can parameterise the chosen analysis in a simple and user-friendly manner on the ROP-R platform.
3. Starting with the chosen module, ROP-R creates for R a proper data file, the necessary R script, runs it, and finally creates an attractive output from the R outputs for the ROP-R platform.
4. In the case of each statistical analysis, ROP-R saves and places the used R-scripts into text files, which may be good for experts in R to check what analysis was actually performed, as well as for learners of R. Advanced R users can modify these scripts, customise them to their own needs, and run them independently of ROP-R on an R platform, such as RGui or R-Studio.

---

[1] See https://www.jamovi.org/features.html

This raises the question of why ROP-R must be used if JASP and jamovi exist with their rich repertoires of statistical analyses. The benefit of ROP-R as an alternative to other competitive statistical software programs may be formulated as follows:

- Those who use ROPstat regularly and are familiar with its structure will probably welcome ROP-R, regarding it as an extension of ROPstat. The user can read and analyse MSW data files of ROPstat directly without any conversion difficulty because ROP-R has the same data-handling facility as ROPstat.
- With ROP-R, one can perform several special statistical analyses, mainly in the area of clustering, which is not currently available in JASP, jamovi, or other well-known menu-driven statistical software.

In the present paper, the general features and technical requirements of ROP-R are briefly outlined, followed by a more detailed summary of its statistical modules. Finally, the paper ends with an illustrative example of a psychological study of adult attachment.

## 1. The general features and the technical requirements of ROP-R

ROP-R is a free multivariate extension of the payable ROPstat without its standard statistical modules. All the information required to use ROPstat can be found in *Vargha et al. (2015)* and www.ropstat.com.

1. ROP-R can be run with the program ROP-R.exe on Windows, where the R engine (specifically the Rcmd program of R) is installed. ROP-R.exe can be downloaded from http://www.bansagi.hu/r/index.html. After successful installation, ROP-R.exe is located in folder „c:_vargha\ropstat". From here on, one can directly run ROP-R or create a shortcut in the taskbar.
2. Because ROP-R runs R-scripts from within ROP-R, for its proper use, the R software is to be installed (https://www.filehorse.com/download-r-for-windows/old-versions/), and the path of program Rcmd.exe is to be specified in the ROP-R menu point Settings/R-path. The suggested version is R-4.1.3, for which the default path is c:\Program Files\R\R-4.1.3\bin\i386\Rcmd.exe. However, path c:\Program Files\R\R-4.1.3\bin\x64\

Rcmd.exe also works. The latter can run 64-bit procedures faster.[2] Newer R versions (4.2.0 and above) may fail to run properly with the ROP-R.

3. In the next step, several R packages must be installed as follows:
    a. R uses a basic program called RGui.exe (in the same folder as Rcmd.exe).
    b. After starting RGui, copy the lines below regularly to the console and press enter. After a couple of minutes, these packages are properly installed, and RGui can be exited.
    ```
    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).
    ```
    c. If the R software with RGui is freshly installed, it is suggested to run only the first line of the above list [install.packages("cluster", dependencies = TRUE)], accept the options offered, and run the other lines (R commands) only after successfully installing the *cluster* package. If any problems occur with the installation of the packages together, try to install them separately.
4. After these steps, ROP-R can be started, and data files can be read in the same way as in ROPstat. The default format is the ROPstat's MSW file format. In addition, ROP-R accepts Excel type XLS, XLSX files[3], tab-delimited and CSV text files, and SPSS SAV and POR files.
5. Having entered a data file, a module can be selected in the Multivariate statistical analyses menu point.

---

[2] See https://askubuntu.com/questions/54296/difference-between-the-i386-download-and-the-amd64

[3] In reading Excel files in ROP-R the worksheet, containing the data table must be set to the active one.

6. During the module run, ROP-R always runs one or more R scripts (with appropriate messages on the screen). After the script is run, the results are written in a text file. ROP-R automatically reads this and creates user-friendly tables for the output that one can send or copy to Excel or Word, respectively.
7. An important feature of ROP-R is that the R scripts run by ROP-R are always written in text files with *.r extension (such as EFA.r, CFA.r, PolReg.r, and MBCA.r), available to the user in a special folder (c:_vargha\ropstat\aktualis). These can be run independently from ROP-R with an appropriate R platform (such as RGui or RStudio), occasionally with special options and modifications required by the user.
8. If the results include visual figures, ROP-R places them in the folder mentioned above and communicates this at the end of the output.

## 2. The statistical menu system of ROP-R

In this section, we present the currently available statistical modules of the ROP-R separately for three types of statistical topics: regression, dimension reduction, and clustering. We mention that other menu points available in ROP-R (File, Edit, Cases, Variables, Transformations, etc.) are logical and evident to use, work exactly the same way as in ROPstat *(Vargha et al., 2015),* and are similar to other software. Therefore, we do not address them below.

### 2.1 The regression modules of ROP-R

ROP-R contains the following three regression modules based on the R packages *jmv,* created for jamovi *(The jamovi project, 2021; Şahin–Aybek, 2019)* and *rlang (Henry–Wickham, 2022):*

- Hierarchical regression,
- Polynomial regression,
- Binary logistic regression (BLR).

Hierarchical regression (HierR)

In this module, each dependent (outcome) variable is regressed on the same set of specified independent (explanatory) variables in a multiple linear regression model *(de Jong, 1999; Tabachnick–Fidell, 2013, Chapter 5).* The independent variables can be sorted into separate blocks using block indices specified by the user. Block indices are integer numbers that may vary between 1 and 9. Variables with the same block index belong to the same block. The main purpose of HierR is to measure and test the effect of a block added to a set of preceding blocks in predicting a dependent variable.

The blocks enter the multiple linear regression model separately in consecutive steps. This is why this type of multiple regression is called hierarchical or sequential regression *(Tabachnick–Fidell, 2013, p. 137).* The final model is the simplest, after which the explained variance ratio ($R^2$) does not increase significantly. For each entering block, the following statistical measures are to be computed:

- Multiple correlation ($R$) of the model containing all independent variables entered.
- The R-squared value ($R^2$) the explained proportion of variance of the cumulative regression model.
- The adjusted $R$-squared value.
- Standard error of regression estimation (RMSE) of cumulative regression model.
- Significance test of the cumulative regression model ($F$-statistic, *df* values, and $p$-value).
- The increase in $R^2$ ($R^2$+) compared to the model that did not include the entering block.
- The significance test of the $R^2$ + increase ($F$ statistic, *df* values, and $p$-value).

The output of the model also contains the basic descriptive statistics of all specified variables for each dependent variable, the regression coefficients (and their significance) of the final model, and the collinearity diagnostics for the independent variables. Concerning collinearity diagnostics, for each independent variable, an $R^2$ value indicates the proportion of variance explained by the other independent variables; TOL = $1 - R^2$, the proportion of variance *not* explained by the other independent variables; and VIF = 1/TOL, the variance inflation factor. Multicollinearity was indicated by very high $R^2$ and VIF and very low TOL values ($R^2 > 0.99$, TOL < 0.01, VIF > 100). Independent variables that cause multicollinearity have to be excluded from the analyses. The presence of multicollinearity may cause an unreliable estimation of the regression coefficients.

A special option in the HierR module is that mediation analysis can be performed if two blocks are defined and the first block does not include more than two variables. In this case, the program asks you if you want mediation analysis to be performed. If the answer is yes, beyond a standard HierR, each independent variable in the second block will be regarded and analysed as a moderator variable in the relationship between the first block variable(s) and each dependent variable, provided that the number of dependent variables is less than 10.

Another option in HierR is the listing of influential cases and saving a variable that measures the influential impact. We consider a data point that is highly influential if removing the point substantially impacts the regression model. In linear regression, Cook's distance (D) is often calculated for each observation to describe the degree of influence of the observation on the model *(Marzjarani, 2015).* Observations with larger D values have a greater influence. Saving D, a robust regression analysis can be performed by omitting highly influential cases by regarding outlier cases above a certain threshold of D and using D as a conditional grouping variable in subsequent ROP-R regression analyses. Typically, any observation with D greater than Dmin = $4/n$ ($n$ = total number of cases) is a potential outlier, although the Dmin threshold is not a rigorous standard. Modifying the default value (=1) of the „multiplier of standard (4/n) threshold" option in the HierR module can either increase or decrease the number of outliers. Checking in the „save influential case variable" option, the D/Dmin variable (Cook`s relative distance) will be saved (attached to the active MSW data file).

After the completion of a HierR analysis, the user will find the data file of the independent and dependent variable (tmpdat.txt), used R-script (HierReg.r), created path plots (or plots) of the mediation analysis (modplot*.pdf), and text file of the R-output (oo.txt) in folder c:_vargha\ropstat\aktualis.

### Polynomial regression (PolR)

In this module, each dependent variable is regressed on each independent variable. The main purpose of PolR is to identify independent variables that have a nonlinear effect on some dependent variables. PolR is basically a HierR analysis, where the powers of the independent variable ($X^1 = X$, $X^2$, $X^3$, …) play the role of the blocks, the consecutive blocks containing the increasing powers of the independent variable (in each block, one power of $X$). A significant effect of a power greater than 1 is evidence of a nonlinear effect. In the regression equation of PolR, $X$ represents the linear component, $X^2$ represents the quadratic component, $X^3$ represents the cubic component, etc. In the module, the user may set the maximal power between 2 and 5.

Variable sets where nonlinear effects occur cannot follow a multivariate normal distribution, the main feature of which is a single dense region of the multivariate data. Therefore, such a sample may be appropriate for exploring several dense regions or value patterns using cluster analysis (*Vargha et al., 2016; Vargha–Bergman, 2019).*

The output of the PolR is very similar to that of the HierR. After the completion of a PolR analysis, the user will find the data file of independent and dependent variables (tmpdat.txt), used R-script (PolReg.r), and text file of the R-output (oo.txt) in folder c:_vargha\ropstat\aktualis.

### Binary logistic regression (BLR)

Logistic regression models for binary response variables allow us to estimate the probability of the outcome (e.g. yes vs. no) based on the values of the independent (explanatory) variables. The BLR module in ROP-R is very similar to that of HierR. The differences are as follows: only dichotomous dependent variables are accepted in BLR, and each is regressed on the same set of specified independent variables. The independent variables can be either continuous (default) or categorical, as decided by the user, and they may be sorted to separate blocks in the same way as in HierR. If an independent variable is set to the categorical type, dummy variables are created based on their varied values.[4]

The blocks consequently enter the regression model separately. The final model is the simplest, after which a newly entered block of independent (explanatory) variables does not significantly increase the fit of the regression model. BLR is basically a multiple linear regression in which the dependent variable is the logit of the relative frequency of the larger value of the dichotomous dependent variable (with the specified independent variables after the necessary dummying), but it can also be regarded as a nonlinear discriminant analysis with a two-group categorical criterion variable *(Tabachnick–Fidell, 2013, Chapter 10).* If $p$ is a number between 0 and 1, then

$$\text{odds}(p) = p/(1-p) \text{ and } \text{logit}(p) = \log(\text{odds}(p)),$$

where log denotes the natural ($e$ based) logarithm.

The output of BLR is also very similar to that of HierR. For each entering block, the following model fit measures are to be computed:

- AIC: Akaike's Information Criterion of the given cumulative BLR model.
- R2-McF: McFadden's pseudo $R^2$ of the given cumulative BLR model.
- R2-CS: Cox & Snell pseudo $R^2$ of the given cumulative BLR model.
- R2-Nag: Nagelkerke's pseudo $R^2$ of the given cumulative BLR model.
- Significance test of the cumulative BLR model ($\chi^2$ statistics, *df* value, *p*-value).

---

[4] For each value and each case, the dummy variable is 1 if the case is characterized with this value and 0 if not.

- Significance test of the improvement of the given BLR model compared to the preceding one ($\chi^2$ statistics, *df* value, *p* value).

The output of the model also contains basic descriptive statistics for all specified variables, and for each predictor, the regression coefficients (and their significance) of the final model. This table also contains the odds ratio = exp(B) value of each *X* predictor and the relative change in the odds per unit increase in the predictor. The relative change is greater than 1 if the expected probability of the larger value of the dependent variable, along with the $p/(1 - p)$ odds, increases owing to the unit increase in *X*, and less than 1 if the expected probability decreases. For categorical predictors, the odds values indicate the relative change in the odds when the value of the predictor changes from the reference value (minimum value) to the value indicated in the current row.

Moreover, the output contains collinearity diagnostics (TOL and VIF measures) for the independent variables (independent variables causing multicollinearity should be dropped from the analysis), classification table for the final model, and summary table of the predictive measures (accuracy, specificity, and sensitivity percentages). The data in the latter two tables depend on the cut-off value of the probability prediction scale, which can be set to any value between 0 and 1 on the panel of the BLR module.

After the completion of a BLR analysis, the user will find in folder c:_vargha\ropstat\aktualis the data file of independent and dependent variables (tmpdat.txt), used R-script (BinLogReg.r), and text file of the R-output (oo.txt).

## 2.2 The dimension reduction modules of ROP-R

ROP-R contains the following three-dimension reduction modules:

- Principal component analysis (PCA),
- Exploratory factor analysis (EFA),
- Confirmatory factor analysis (CFA).

Principal component analysis (PCA)

This module performs a standard principal component analysis of quantitative variables (*Tabachnick–Fidell, 2013, Chapter 13).* The extracted principal components can be rotated using an orthogonal (Varimax) or oblique (Promax) method. The PCA output contains the following standard tables:

- Basic descriptive statistics.
- Eigenvalues and explained variances of principal components.

- Matrix of principal component loadings (correlations between variables and components).
- Rotated matrix of principal component loadings (if rotation is requested).
- Correlation matrix of rotated components (if oblique rotation is requested).
- Structure matrix, which is the table of correlations between variables and rotated components (if oblique rotation is requested).

A special option of this module (based on R package *MBESS*; *Kelley, 2007)* is that one can request internal consistency reliability measures of the input variables. Choosing this option, the program computes Cronbach's alpha and McDonald's omega reliability measures for the input variables, regarding them as items of a single scale. It is important to know if an item is to be reversed, this transformation must be performed on the item prior to reliability analysis. Within the ROP-R, this can be done by using the Transformations/Univariate function consecutively with the cX menu point (setting $c = -1$) and the X+c menu point (setting $c$ = item minimum + item maximum) or using the item conversion option in the *item analysis* module of ROPstat.

Another option in PCA is listing and saving outlier cases (similar to influential cases in HierR) using the R package MASS *(Venables–Ripley, 2002).* In this approach, the detection of multivariate outliers is based on a robust variant of the Mahalanobis distance *(Leys et al., 2018)* and the distance of each case from the centre of the total sample; as a measure of case extremity, RMD can be saved as a new variable. By means of this variable, which plays a role similar to Cook's distance in regression analyses (see Section 2.1), robust principal component and factor analyses can be performed without the influence of outliers. Checking in the „save variable of case extremity" option in the PCA module, the RMD variable will be saved (attached to the active MSW data file). ROP-R uses a special threshold to identify outliers that are often less restrictive (yielding fewer outliers) than the built-in threshold of Cook's distance in HierR. However, decreasing the default value (=1) of the „Multiplier of built-in threshold" option in the PCA module can increase the number of outliers.

After the completion of the PCA analysis, the user will find the data file of the input variables (tmpdat.txt) and used R-scripts (PCA.r, rotate.r) in folder c:_vargha\ropstat\aktualis.

### Exploratory factor analysis (EFA)

Based on R packages *psych (Revelle, 2022), olsrr (Hebbali, 2020),* and *GPArotation (Bernaards–Jennrich, 2005),* this module performs exploratory factor analysis of quantitative variables with three different optional methods: maximum likelihood (ML), principal axis factoring (PAF), and minimum residual (MinRes) *(Osborn, 2014; Tabachnick–Fidell, 2013, Chapter 13).* The extracted primary factor structure can be rotated using an orthogonal (Varimax) or oblique (Promax) method. The output of the EFA contains the following standard results.

- Basic descriptive statistics.
- Kaiser–Meyer–Olkin adequacy measure (KMO).
- Eigenvalues and explained variances of principal components.
- Matrix of principal component loadings (correlations between variables and components).
- Collinearity diagnostics of the input variables ($R^2$, TOL, and VIF measures) − variables causing multicollinearity should be dropped from the analysis.
- Test of factor model fit.
- Some adequacy measures of the explored model (RMSEA, RMSR, CFI, and TLI).
- Pattern matrix, which is the matrix of rotated factor loadings (if rotation is requested).
- Correlation matrix of rotated factors (if oblique rotation is requested).
- Structure matrix, which is the table of correlations between variables and latent factors (if oblique rotation is requested).

After the completion of an EFA analysis, the user will find the data file of the input variables (tmpdat.txt), and used R scripts (Rbegin.r, EFA.r) in folder c:_vargha\ropstat\aktualis.

### Confirmatory factor analysis (CFA)

Based on R packages *lavaan (Rosseel, 2012)* and *lavaanPlot (Lishinski, 2021),* this module performs a confirmatory factor analysis (CFA) of factor models defined by theoretical considerations or explored on other independent samples *(Harrington, 2009; Rosseel et al., 2018).* The input variables (items) can be sorted into separate factors (scales) using factor indices specified by the user. These indices are integer numbers that may vary between 1 and 9. Variables with the same factor index belong to the same factor. The main purpose of CFA is to test the specified factor model (via $\chi^2$ tests) and measure the goodness of the model fit using several adequacy measures.

The specified model can be estimated using three different basic methods (ML = maximum likelihood, ULS = unweighted least squares, DWLS = diagonally weighted least squares) and their five robust versions:

- MLMV: ML estimation with robust standard errors and a mean- and variance adjusted test statistic.

- MLR: ML estimation with Huber–White robust standard errors and a scaled test statistic that is asymptotically equal to the Yuan–Bentler test statistic.
- ULSMV: Robust variant of ULS estimation with robust standard errors and a mean and variance-adjusted test statistic.
- WLSM: DWLS estimation with robust standard errors and a mean adjusted test statistic.
- WLSMV: DWLS estimation with robust standard errors and mean and variance-adjusted test statistic.

CFA runs with the chosen estimator of the specified model and an improved model if one or more modification indices of the first run exceed the specified limit for the within- or between-factor residuals, or for the covariances of cross-loadings. The CFA output contains the following standard results:

- Basic descriptive statistics.
- The modification indices of covariances of the specified model.
- $\chi^2$ tests of the specified and baseline models.
- Fit indices for the different methods and models (AIC, BIC, RMSEA, pClose, CFI, TLI, and SRMR).
- Standardized factor loadings and communalities of the specified model.
- Pairwise standardized covariance (= correlation) estimates of the latent factors.
- The largest modification indices of covariances in the improved model (if one or more modification indices of the first run exceed the specified limit).
- Table of standardised factor loadings of latent factors, including cross-loadings for the improved model (if any modification index of the covariances of cross-loadings of the first run exceeds the specified limit).
- Table of estimations of standardised residual covariances (correlations) for the improved model (if any modification index of within- or between-factor covariances exceeds the specified limit).

After the completion of a CFA analysis, the user will find the data file of the input variables (tmpdat.txt), used R Rscripts (CFA.r and CFA2.r), text files of the R output (oo.txt and o2.txt), and PDF files of the prepared path diagrams (pathplot1.pdf and pathplotR1.pdf) in folder c:_vargha\ropstat\aktualis. If a conditional group is used, path diagram will be created for each group.

## 2.3 The clustering modules of ROP-R

ROP-R offers a rich menu of clustering procedures *(Bergman et al., 2003; Kaufman–Rousseeuw, 2009),* many of which are unavailable in other user-friendly statistical software. ROP-R contains the following four clustering modules:

- Agglomerative hierarchical cluster analysis (AHCA),
- Divisive hierarchical cluster analysis (DHCA),
- *k*-center cluster analysis (KCA),
- Model-based cluster analysis (MBCA).

Both AHCA and DHCA produce a hierarchical series of clustering of a given sample of observations based on some (not too many) quantitative input variables, best illustrated by a dendrogram *(Roux, 2018).* In the first step of AHCA, each observation is regarded as a one-member cluster, and in each step, the two most similar clusters are fused into one (agglomerative clustering algorithm). In contrast, in the first step of DHCA, the whole sample is regarded as one big cluster containing all observations, and in each step, the most heterogeneous cluster is divided into two sub-clusters (divisive clustering algorithm).

In non-hierarchical KCA, one tries to partition a sample with a *k* pre-defined cluster number, *k* optimally homogeneous, and well separable clusters *(Kaufman–Rousseeuw, 2009, Chapter 2).* In MBCA, it is assumed that the multivariate data come from a mixture of different subpopulations following given distributions, typically a multivariate normal. In this framework, MBCA is the exploration of the underlying mixture structure, deciding on size (cluster number) and type of cluster structure *(Fraley–Raftery, 2002; Gergely–Vargha, 2021).*

Despite these possibilities, ROP-R does not cover all clustering methods available in ROPstat, which is specially designed for pattern-oriented analyses *(Vargha et al., 2015; Vargha et al., 2016).* In this respect, ROP-R can be regarded as an extension of ROPstat, which is suggested for use together with ROPstat.

### Agglomerative hierarchical cluster analysis (AHCA)

Based on R packages *stats (R Core Team, 2021)* and *cluster (Maechler et al., 2022),* this module performs standard agglomerative hierarchical cluster analysis *(Roux, 2018)* using six optical distance types (Squared Euclidean, Euclidean, Manhattan, Canberra, Maximum, Minkowski) and eight agglomerative methods (Average, Single, Complete, Centroid, Median, Ward, Flexible beta, and McQuitty). Four optional diagrams (dendrogram, Silhouette plot, Total WSS plot, and Banner diagram) may help evaluate the results.

The user may choose the cluster numbers for which the following results are provided.

- Three adequacy measures (HCmean, EESS% = explained error sum of squares %, XBmod = modified Xie-Beni index; see *Vargha et al., 2016)* of the cluster structure.
- Cluster statistics.
- Pattern of standardized means.

If requested, cluster code variables can be saved (attached to the end of the actual MSW data file).

After the completion of AHCA, the user will find the data file of the input variables (tmpdat.txt), data file extended with cluster code variables for the requested cluster numbers (tmpdat2.txt), used R script (AHCA.r), and requested diagrams in jpg or PDF files (e.g., Dendr1.jpg or Banner1.pdf) in folder c:_vargha\ropstat\aktualis. Diagrams are created for each group if a conditional group is used.

Divisive hierarchical cluster analysis (DHCA)

Based on R packages *cluster (Maechler et al., 2022), factoextra (Kassambara–Mundt, 2020),* and *ggplot2 (Wickham, 2016),* this module performs DIANA divisive hierarchical cluster analysis (*Kaufman–Rousseeuw, 2009, Chapter 6)* using six optional distance types (Squared Euclidean, Euclidean, Manhattan, Canberra, Maximum, Minkowski). Optional diagrams, output, and saving options are the same as those in AHCA.

After the completion of DHCA, the user will find the data file of the input variables (tmpdat.txt), data file extended with cluster code variables for the requested cluster numbers (tmpdat2.txt), used R script (DHCA.r), and requested diagrams in JPG or PDF files in folder c:_vargha\ropstat\aktualis.

*k*-center cluster analysis (KCA)

Based on R packages *stats (R Core Team, 2021), cluster*s (*Maechler et al., 2022), ClusterR (Mouselimis, 2022), Gmedian (Cardot, 2022), factoextra (Kassambara–Mundt, 2020),* and *ggplot2 (Wickham, 2016),* this module can perform three types of *k-centre* cluster analysis: *k*-means, *k*-medoids, and *k*-medians. The latter two are suggested when the input variables are seriously non-normal or ordinal *(Kaufman–Rousseeuw, 2009, Chapter 2).* The best-known *k*-means analysis can be performed with three optional algorithms (Hartigan–Wong, MacQueen, and Lloyd–Forgy). Several types of plots (Silhouette, EESS%, mean heterogeneity, and f(K) distortion) help determine the optimal cluster number. The structures of the output and saving options are similar to those of AHCA.

After the completion of KCA, the user will find the data file of the input variables (tmpdat.txt), data file extended with the cluster code variable for the specified cluster number (tmpdat2.txt), used R script (KCA.r), and requested diagrams in JPG files in folder c:_vargha\ropstat\aktualis.

Model-based cluster analysis (MBCA)

Based on R packages *mclust* (*Scrucca et al., 2016), factoextra (Kassambara–Mundt, 2020),* and *ggplot2 (Wickham, 2016),* this module can perform a model-based cluster analysis. The program searches for each model belonging to different cluster numbers (specifiable between 2 and 25) and to a maximum of 14 possible model types, the best fit by means of an ML algorithm *(Fraley–Raftery, 2002).* The best model has the largest BIC or ICL criterion value.

The evaluation of the results was based on the BIC or ICL plot and summary tables of the optimal solution. The structures of the output and saving options are similar to those in KCA. Options in MBCA are the tables of BIC and ICL values, the clustering *p*-values, and the saving of the uncertainty values in a new variable of the data file.

After the completion of MBCA, the user will find the data file of the input variables (tmpdat.txt), data file extended with the cluster code variable for the optimal MBCA solution (tmpdat2.txt), used R script (MBCA.r), and requested diagrams (e. g., BIC plot) in JPG files in folder c:_vargha\ropstat\aktualis.

## 3. An illustrative example using ROP-R

In a study on psychological attachment, we obtained data on the Experiences in Close Relationships – Relationship Structures (ECR-RS) questionnaire *(Fraley et al., 2011)* from 336 Hungarian adults (124 males and 212 females) living in hetero relationships[5] *(Jantek–Vargha, 2016).* This questionnaire assesses individual differences in attachment within and across various relational contexts, including attachment to mothers, fathers, romantic partners, and friends. In the 40-item self-report questionnaire, 10 items (6 measuring avoidance and 4 measuring anxiety) belonged to each of the four domains with the same questions. In each domain, the Anxiety and Avoidance subscales were

---

[5] Many thanks to Gyöngyvér Jantek who agreed for her data sample to be used in the statistical analyses presented in this paper.

formed from the corresponding items. In the attachment model, anxiety and avoidance were the two main components. Persons with no attachment problems have a low level of anxiety and avoidance, and attachment is highly problematic if anxiety and avoidance demonstrate high level *(Fraley et al., 2011).*

In the following sections, we present some steps of the Hungarian adaptation of the ECR-RS using several modules of ROP-R. For illustration purposes, we chose the Partner scale with its 10 items: AvP01, AvP02, AvP03, AvP04, AvP05, AvP06, AnxP07, AnxP08, AnxP09, and AnxP10 (where Av stands for avoidance, Anx for anxiety, and P for partner).

1. In the first step, we performed PCA on the above 10 items using the PCA module of ROP-R. We checked the reliability measure options and obtained the following results:

- Cronbach's alpha = .875, CI_.95 = (.855; .889),
- McDonald's omega = .871, CI_.95 = (.834; .909).

As the reliability estimations are well above .80, these values confirm the internal consistency of the Partner attachment scale of the ECR-RS *(DeVellis, 2016).*

In the PCA, the first three eigenvalues exceeded 1 ($\lambda_1 = 4.79$, $\lambda_2 = 1.41$, $\lambda_3 = 1.03$, $\lambda_4 = 0.59$), accounting for 72.4% of the total variance, indicating a three-factor structure of the 10 items.

2. In the next step, we performed an EFA on the above 10 items using the EFA module of ROP-R. Since the 7-point items were all strongly non-normal (either skewness or kurtosis, or both were significant at *the* $p < 0.001$ level for all 10 items), the principal axis factoring method of EFA was chosen with Promax oblique rotation of the first three factors. The rotated factor loadings are listed in Table 1.

Table 1

**Sorted rotated factor loading matrix of the 10 items of the Partner scale***

| Item | Factor1 | Factor2 | Factor3 |
|---|---|---|---|
| AvP02 | 0.898 | | |
| AvP03 | 0.832 | | |
| AvP01 | 0.765 | | |
| AvP04 | 0.601 | | |
| AnxP09 | | 0.913 | |
| AnxP07 | | 0.831 | |
| AnxP08 | | 0.690 | |
| AnxP10 | | 0.575 | |
| AvP05 | | | 1.003 |
| AvP06 | | | 0.591 |

* Loadings under .2 are not presented.

Table 1 shows that Avoidance and Anxiety items clearly form different factors. An astonishing result is that two items (AvP05 and AvP06) of the avoidance subscale form a different factor, independent of the other avoidance items. This probably happened because these two items were formulated in a negative form (for this reason, they had been converted prior to PCA and EFA). Therefore, we omitted items AvP05 and AvP06 from further analyses and defined subscales Avoidance and Anxiety based on the items of Factor1 and Factor2 in Table 1. The repeated EFA performed on the retained 8 items of the Partner scale had an excellent KMO value of 0.835, and the rotated factor loading matrix (see Table 2) confirmed the existence of the Avoidance and Anxiety subscales.

Table 2

**Sorted rotated factor loading matrix of the 8 positive items of the Partner scale***

| Item | Factor1 | Factor2 |
|---|---|---|
| AvP02 | 0.932 | |
| AvP03 | 0.819 | |
| AvP01 | 0.722 | |
| AvP04 | 0.630 | |
| AnxP09 | | 0.928 |
| AnxP07 | | 0.846 |
| AnxP08 | | 0.648 |
| AnxP10 | | 0.588 |

* Loadings under .2 are not presented.

3. We subsequently performed a CFA analysis on the Partner scale with the retained 8 items, applying the robust MLMV method. We chose this method among the five available robust methods in the CFA module of ROP-R because earlier studies indicated that MLMV is a reliable and efficient CFA model fitting method (*Gao et al., 2020; Zábó et al., 2022).* In the first run, two *c* modification indices exceeded 20 (between AnxP07 and AnxP09 c = 22.69, between AvP02 and AvP03 c = 20.29). After building these covariances into the factor model (setting the modification index thresholds to 20 for each covariance type), our model could not be significantly rejected ($p > 0.35$, see the last column of Table 3). This model had excellent fit indices (see the last column of Table 4) because the $\chi^2/df$ value (parsimony fit index) was less than 3.5, RMSEA and RSMR were less than 0.05, pClose > 0.50, and CFI, TLI > 0.95 *(Hu–Bentler, 1999; Browne–Cudeck, 1993; Kline, 2005).*

Table 3

**Testing model fit***

| Model | Standard ML | Robust MLMV | Improved ML | Improved MLMV |
|---|---|---|---|---|
| $\chi^2$ statistic | 65.58 | 40.35 | 26.9 | 18.5 |
| *df* | 19 | 19 | 17 | 17 |
| *p* value | < 0.001 | 0.003 | 0.060 | 0.358 |

* The improved models include covariances based on large modification indices.

Table 4

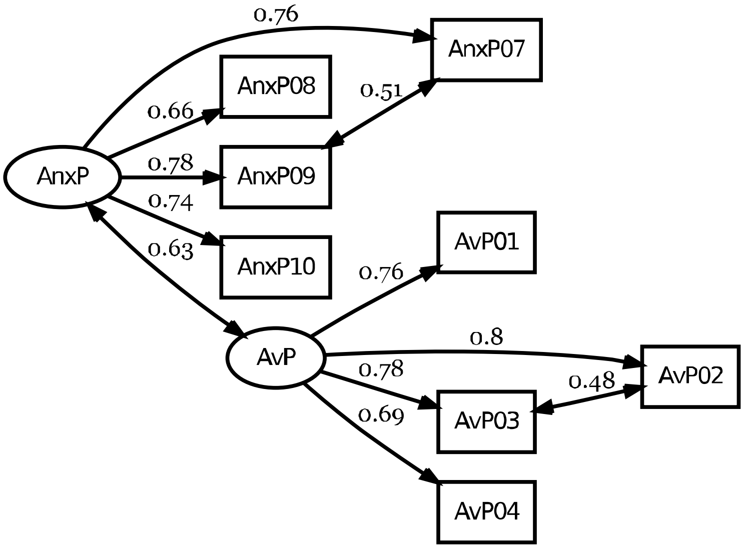
**Model fit indices***

| Model | Standard ML | Robust MLMV | Improved ML | Improved MLMV |
|---|---|---|---|---|
| $\chi^2/df$ | 3.45 | 2.12 | 1.58 | 1.09 |
| RMSEA | 0.086 | 0.058 | 0.042 | 0.016 |
| CI_.95 (RMSEA) | (0.064; 0.110) | (0.033; 0.084) | (0.000; 0.071) | (0.000; 0.054) |
| pClose | 0.005 | 0.264 | 0.638 | 0.923 |
| CFI | 0.966 | 0.966 | 0.993 | 0.998 |
| TLI | 0.950 | 0.950 | 0.988 | 0.996 |
| SRMR | 0.044 | 0.044 | 0.029 | 0.029 |

* The improved models include covariances based on large modification indices.

Figure 1

**The path diagram with the standardised regression estimates of the improved MLMV model of the Partner scale with two subscales: Anxiety (AnxP) and Avoidance (AvP)**



The path diagram (with the standardised regression estimates) of the improved MLMV model of the Partner scale with the two subscales, Anxiety (AnxP) and Avoidance (AvP), is shown in Figure 1. ROP-R created a PDF file (named pathplotR1.pdf), which can be converted free of charge to a JPG file using the website https://pdf2jpg.net/.

The overall results obtained confirm that the Hungarian adaptation of the ECR-RS has good structural validity. Additional evidence for simplifying the model of the Partner scale of ECR-RS by omitting items AvP05 and AvP06 was that performing a CFA with all 10 original Partner items, AvP05 and AvP06 had the smallest standardised loadings (less than .30) in the improved MLMV model, whereas all the other standardised loadings were greater than 0.45. Having these good results, we created the Anxiety_P and Avoidance_P subscales of the Partner scale by averaging the appropriate items, using a special option of module CFA, to perform further statistical analyses with these subscales.

4. In addition to the structural reliability, it is important to prove the validity of an adapted scale. Several studies have demonstrated that adult attachment and mental health are positively correlated. Individuals with a disordered attachment style (with high levels of anxiety and avoidance) usually have a lower level of mental health *(e.g., Adams et al., 2018).* This was also confirmed in our study, where the level of mental health was assessed using the 5-item WHO Well-Being Scale (WBI-5; *Bech, 1996, 2012).* Using module HierR of ROP-R with subscales Anxiety_P and Avoidance_P as independent variables, and WBI-5 as dependent variable, the $R^2$ explained variance of the two predictors was 0.122 ($F(2; 321) = 22.28, p < 0.001$).

However, the question arises whether this influence of attachment still remains if the basic personality traits of the Big Five model enter the regression model first. In our study, the Big Five model was measured using five scales (Extraversion, Agreeableness, Neuroticism, Conscientiousness, Openness) from the BFI-44 Big Five Inventory *(John et al., 2008).* Performing a hierarchical regression analysis using module HierR of ROP-R, the first block was defined using the five BFI-44 scales and the second block with the two attachment scales (Anxiety_P and Avoidance_P). The results showed that the first block with the five BFI-44 scales explained 27.9% of the variance of WBI-5, but Anxiety and Avoidance, entering in a second block could increase this value by a significant ($F(2; 316) = 11.21, p < 0.001$) and explainable amount of 0.048 to a level of $R^2 = 0.327 = 32.7\%$.

5. Having a reliable and meaningful Partner attachment subscale, we explored basic attachment types using them as input variables in cluster analyses. If the input variables have a multivariate normal distribution with only one single centre, no different types can be expected. To exclude this possibility, we

performed polynomial regression analyses between Anxiety_P and Avoidance_P, where strong nonlinear relationships contradicted multivariate normality.

Using Anxiety_P as the dependent variable and Avoidance_P as a predictor in the PolR module of ROP-R, we found significant nonlinear effects (see Table 5), which disproves the bivariate normality of the two subscales and provides the possibility to explore real attachment types by means of cluster analysis.

Table 5

**Polynomial regression of Anxiety_P on predictor Avoidance_P**

| Model | $R^2$ | $R^2$ increase | $F$ | $df1$ | $df2$ | $p$-value |
|---|---|---|---|---|---|---|
| Linear | 0.258 | | | | | |
| Quadratic | 0.268 | 0.010 | 4.401 | 1 | 327 | 0.037 |
| Cubic | 0.280 | 0.012 | 5.493 | 1 | 326 | 0.020 |
| 4th grade | 0.285 | 0.005 | 2.218 | 1 | 325 | 0.137 |
| 5th grade | 0.306 | 0.021 | 9.733 | 1 | 324 | 0.002 |

6. Among the cluster analyses, we first performed a standard Ward type agglomerative hierarchical analysis using the AHCA module of ROP-R, standardising the input variables. We asked for a dendrogram (see Figure 2), Silhouette plot (see Figure 3), and detailed results for clusters 3 to 6. Based on these results, the adequacy measures are summarised in Table 6. For each cluster, the homogeneity coefficient (HC) is the mean within-cluster distance with the average squared Euclidean distance. The smaller the HC, the more homogeneous the cluster. HCmean is the average of the HC values (weighted by the cluster sizes) in a cluster solution *(Vargha et al., 2016),* where HCmin and HCmax denote the minimum and maximum HC values, respectively. EESS%, which is the explained error sum of squares percentage, is also a cohesion measure of a cluster structure. EESS% values above 65 indicate good structural integrity. XBmod, the modified Xie-Beni index, similar to the Silhouette index, indicates how well the clusters separate from each other in a cluster solution. For both indices, values above 0.50 indicate a good structure *(Vargha et al., 2016; Bergman et al., 2017).*

In AHCA, the dendrogram (see Figure 2) does not indicate the optimal cluster number. However, the Silhouette plot indicates that hierarchical cluster solutions for $k = 2$ to 4 have an acceptable level of the Silhouette index (see Figure 3). The results summarised in Table 6 indicate that EESS% only has an acceptable level above $k = 3$, and for the $k = 4$ solution, HCmean and XBmod also have good values. Only HCmax was extremely high, indicating that the cluster with the largest HC value was highly heterogeneous. Going with $k$ a little upward (until $k = 6$), HCmax never decreased. For $k = 5$, the surprisingly low value of XBmod

(= 0.094) may be due to the fact that in this solution there are two highly similar clusters, the fusion of which in the *k* = 4 solution already leads to a good level of cluster separability for both XBmod (see Table 6) and the Silhouette index (see Figure 3).

Figure 2

**The dendrogram of Ward-type AHCA analysis with input variables Anxiety_P and Avoidance_P**

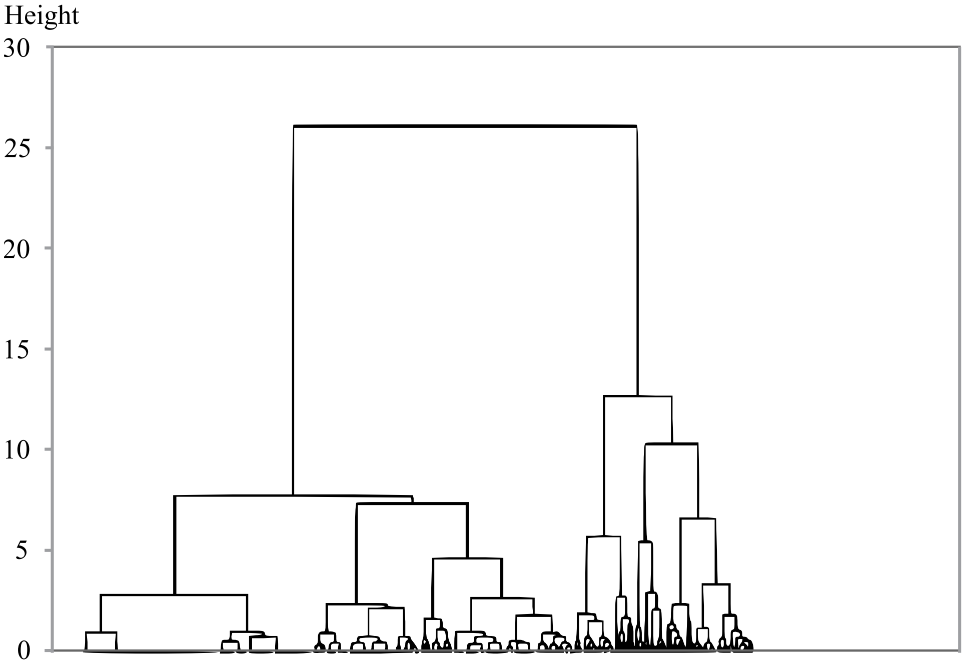


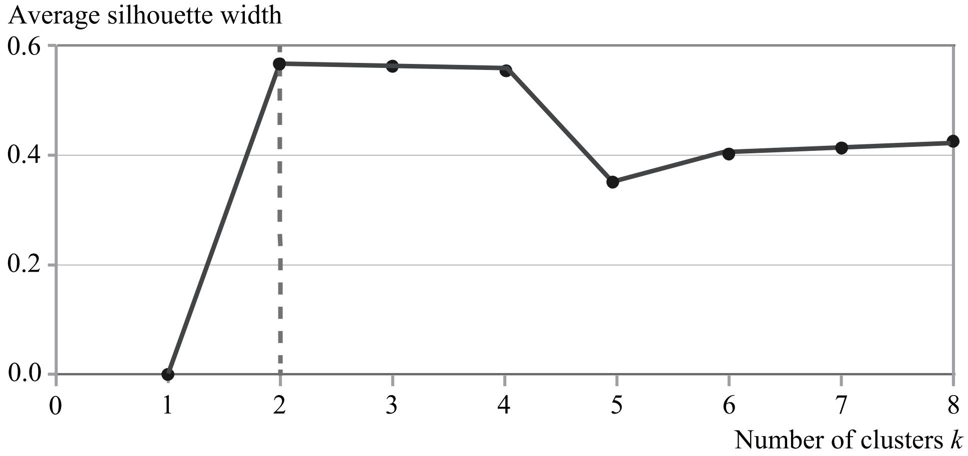
Table 6

**Adequacy measures of AHCA analyses with Anxiety_P and Avoidance_P for cluster numbers 3 to 6**

| Cluster number | EESS% | HCmean | HCmin | HCmax | XBmod |
|---|---|---|---|---|---|
| *k* = 6 | 81.234 | 0.386 | 0.050 | 1.553 | 0.451 |
| *k* = 5 | 77.068 | 0.469 | 0.050 | 1.553 | 0.094 |
| *k* = 4 | 72.448 | 0.561 | 0.382 | 1.553 | 0.878 |
| *k* = 3 | 64.309 | 0.722 | 0.382 | 1.983 | 0.829 |

Figure 3

**The Silhouette plot of Ward-type AHCA analysis with Anxiety_P and Avoidance_P**



Based on the detailed AHCA results, the $k = 4$ solution appears to be the most promising. On the ROP-R output, we find a table of standardised means of the explored clusters (see Table 7) that helps explain the obtained cluster solution. Cluster CL1 is the largest, with its moderately low anxiety – moderately low avoidance pattern, it represents the most homogeneous (HC = 0.38) and most common ($n = 242 = 73.3\%$) good attachment type. The peculiarity of CL2 is its very high avoidance, the peculiarity of CL3 is its very high anxiety, and CL4 represents the most problematic attachment type with very high levels of both anxiety and avoidance. Although this latter cluster, CL4, seems to be very heterogeneous (HC = 1.55), the obtained structure matches the theoretical model of attachment and is in line with earlier findings (*Fraley et al.*, 2011).

Table 7

**Standardized means of the k = 4 AHCA solution with the HC values of the clusters**

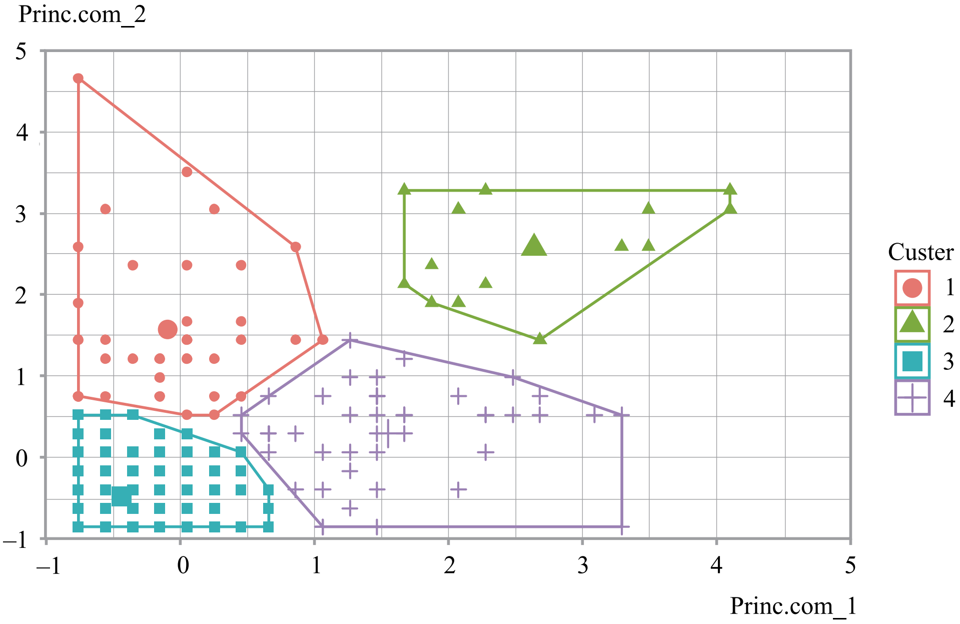
| Cluster | Anxiety_P | Avoidance_P | CLsize | HC |
|---|---|---|---|---|
| CL1 | −0.41 | −0.46 | 242 | 0.38 |
| CL2 | −0.13 | 1.82 | 29 | 0.99 |
| CL3 | 1.50 | 0.46 | 43 | 0.91 |
| CL4 | 2.34 | 2.43 | 16 | 1.55 |

7. Among further cluster analyses, the DHCA module of ROP-R yielded cluster solutions that were more difficult to explain than the four-cluster AHCA solution. Model-based cluster analysis (performed with the MBCA module of ROP-R) did not yield a clear and well-explainable solution.

By performing *k-centre* analyses with the KCA module of ROP-R, the *k*-means method yielded the best solution again with four clusters. This had somewhat better adequacy measures (EESS% = 75.5, HCmean = 0.50, XBmod = 0.887) than the four-cluster AHCA solution but had a very similar pattern (see Figure 4 below, created by means of the „plot obtained clusters" option of the KCA module of ROP-R). In Figure 4, the four clusters are placed in the two-dimensional space of the first two principal components of Anxiety_P and Avoidance_P. Cluster 3 matches well with the highly homogeneous CL1 cluster in the four-cluster AHCA solution, cluster 2 is the highly extreme CL4 cluster, and clusters 1 and 4 are the remaining two clusters of the four-cluster AHCA solution.

Figure 4

**The four-cluster solution of the k-means analysis with Anxiety_P and Avoidance_P**



# 4. Discussion

In this study, we introduced ROP-R, a free ROPstat-based multivariate statistical program package that can be run on a Windows system. ROP-R has a user-friendly menu system for running several statistical procedures (regressions, dimension reduction analyses, and clustering procedures) using ROP-R-created R scripts, similar to some other recently developed software such as JASP or jamovi. However, the ROP-R modules offer several new and comfortable opportunities that are not available in other software. The usefulness and elegance of the ROP-R were illustrated by a complex series of statistical analyses conducted on a sample of a study on the psychological topic of attachment.

A standard installation of ROP-R can be done in the following steps.

- Install ROP-R from the website http://www.bansagi.hu/r/index.html. After a successful installation, ROP-R.exe is found in folder c:_vargha\ropstat.
- Install R version R-4.1.3 for Windows (https://www.filehorse.com /download-r-for-windows/old-versions/ (if not installed on your computer yet). Newer R versions (4.2.0 and above) may fail to properly run with the ROP-R.
- Start ROP-R.exe and specify the path of program Rcmd.exe in the menu point Settings/R-path (the standard path is c:\Program Files\R\R-4.1.3 \bin\x64\Rcmd.exe).
- Run RGui.exe (if there is no desktop shortcut, the standard folder of RGui.exe is c:\Program Files\R\R-4.1.3\bin\x64).
- If R packages have been installed earlier on your PC, install all packages in RGui listed at point 3. b in Section 1 at the beginning of this paper.
- If you are new in R and no R packages have been installed earlier on your PC, copy the command „install.packages("cluster", dependencies = TRUE)" to RGui, press enter, and accept the offered options. If you receive positive feedback on the installation of the package *cluster*, copy the remaining lines (R commands) from Point 3. b of Section 1 at the beginning of the present paper and press enter.

The aforementioned steps were performed only once. After completing them, you can enjoy running several modules of the ROP-R, detailed in Section 2 of our paper.

## References

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.* Vol. 63. No. 10. pp. 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.

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, London.

Bergman, L. R. – Vargha, A. – Kövi, Z. (2017): Revitalizing the typological approach: Some methods for finding types. *Journal for Person-Oriented Research*. Vol. 3. No. 1. pp. 49–62. DOI: 10.17505/jpor.2017.04; https://www.person-research.org/journal/files/3_1/filer/4.pdf

Bernaards, C. – Jennrich, R. I. (2005): Gradient Projection Algorithms and Software for Arbitrary Rotation Criteria in Factor Analysis. *Educational and Psychological Measurement*. Vol. 65. No. 5. pp. 676–696.

Browne, M. W. – Cudeck, R. (1993): Alternative ways of assessing model fit. In: Bollen, K. A. – Long, J. S. (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.* Vol. 97. No. 458. pp. 611–631. https://doi.org/10.1198/016214502760047131

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. *Psyhological Assessment.* Vol. 23. No. 3. pp. 615–625.

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.* Vol. 27. No. 2. pp. 192–201. DOI: 10.1080/10705511.2019.1637741

Gergely, B. – Vargha, A. (2021): How to Use Model-Based Cluster Analysis Efficiently in Person-Oriented Research. *Journal for Person-Oriented Research*. Vol. 7. No. 1. pp. 22–35. 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. – P. M. Bentler (1999): Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*. Vol. 6. No. 1. pp. 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.* Vol. 71. No. 3. pp. 447–470. 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. – Naumann, L. P. – Soto, C. J. (2008): Paradigm Shift to the Integrative Big-Five Trait Taxonomy: History, Measurement, and Conceptual Issues. In: John, O. P. – Robins, R. W. – Pervin, L. A.: *Handbook of personality: Theory and research.* pp. 114–158. New York: Guilford Press.

de Jong, P. F. (1999): Hierarchical regression analysis in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*. Vol. 6. No. 2. pp. 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

Kline, R. B. (2005): *Principles and practice of structural equation modeling (2$^{nd}$ ed.).* New York, NY: Guilford Press.

Leys, C. – Klein, O. – Dominicy, Y. – Ley, C. (2018): Detecting multivariate outliers: Use a robust variant of Mahalanobis distance. *Journal of Experimental Social Psychology.* Vol. 74. pp. 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*. Vol. 4. No. 2. pp. 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.* pp. 8–181. Victoria, BC, Canada: Leanpub. 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

Rosseel, Y. (2012): lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software.* Vol. 48. No. 2. pp. 1–36. https://doi.org/10.18637/jss.v048.i02

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

Roux, M. (2018): A comparative study of divisive and agglomerative hierarchical clustering algorithms. *Journal of Classification*. Vol. 35. No. 2. pp. 345–366.

Şahin, M. – Aybek, E. (2019): Jamovi: an easy-to-use statistical software for the social scientists. *International Journal of Assessment Tools in Education*. Vol. 6. No. 4. pp. 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.

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

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

Vargha, A. – Torma, B. – Bergman, L. R. (2015): ROPstat: A General Statistical Package Useful for Conducting Person-Oriented Analyses. *Journal for Person-Oriented Research*. Vol. 1. No. 1–2. pp. 87–98. DOI: 10.17505/jpor.2015.09.; https://www.person-research.org/journal/files/1_1/filer/44.pdf

Vargha, A. – Bergman, L. R. (2019): MORI coefficients as indicators of 'real' cluster structure. *Hungarian Statistical Review*. Vol. 2. No. 1. pp. 3–23. DOI: 10.35618/hsr2019.01.en003.

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*. Vol. 2. No. 1–2. pp. 78–86. DOI: 10.17505/jpor.2016.08. https://www.person-research.org/journal/files/2_1/filer/22.pdf

Venables, W. N. – Ripley, B. D. (2002): *Modern Applied Statistics with S. Fourth Edition*. Springer, New York.

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