

ROP-R: a free multivariate statistical software that runs R packages in a ROPstat framework¹

András Vargha

Professor Emeritus, Károli Gáspár University of the Reformed Church,

Professor Emeritus, Eötvös Loránd University,

Budapest, Hungary

E-mail: vargha.andras@kre.hu

Péter Bánsági

Budapest, Hungary

E-mail: peter@bansagi.hu

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 a separate chapter 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

ROPstat is one of the few Hungarian developments of wide scope statistical softwares that has attained some international reputation (Vargha *et al.*, 2015). ROPstat offers specialties in robust techniques, ordinal analyses, and pattern and person oriented methods, many of them not available in other common statistical softwares. In spite of all its merits, ROPstat users may feel bothering that the software lacks basic multivariate methods such as nonlinear regression or dimension reduction methods (e.g., principal component analysis and factor analysis).

Several decades ago, paid softwares (SPSS, SAS, etc.) dominated the statistical software market. However, some twenty years ago, the appearance of R, an open source free software started to change the mind of software developers and users. By now, almost all types of statistical methods can be performed by means of one or more R packages. One difficulty with R is that any analysis in it can only be performed by precisely prepared set of R commands, called scripts, satisfying strict syntax rules. For example, lower and upper case letters are seldom interchangeable in R commands.

To overcome this difficulty, two user-friendly statistical softwares, namely JASP (JASP Team, 2012) 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 which is intuitive to use, and can provide the latest developments in statistical methodology. Their common philosophy is that scientific software should be community driven, where anyone

¹ 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). Accepted for publication in *Hungarian Statistical Review*

can develop and publish analyses, and make them available freely to a wide audience². Both softwares are based on R, but provide a user-friendly menu system for a large scale of statistical analyses without the need of 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 be a ROPstat based free multivariate statistical program package; 2) ROP-R provide 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 ROP-R has been 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. ROP-R is a free, bilingual (English and Hungarian) multivariate statistical software which provides analyses – in its present version – in ten statistical modules of three types:
 - a. Regression
 - b. Dimension reduction
 - c. Clustering.
2. In each module, the user can parametrize the chosen analysis in a simple and user-friendly way on the ROP-R platform.
3. Starting the chosen module, ROP-R creates for R a proper data file, the necessary R-script, runs it, and finally creates from the R outputs an attractive output 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 that may be useful for experts in R for checking what analysis was actually performed, as well as for learners of R. Advanced R users can even modify these scripts, customizing them to their own needs, and can run them independently of ROP-R in an R platform like RGui or R-Studio.

The question may be raised, why to use ROP-R if there are JASP and jamovi with their rich repertoires of statistical analyses? The benefit of ROP-R as an occasional alternative to other competitive statistical softwares 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. Since ROP-R has the same data handling facility as ROPstat, the user can read-in and analyze *.msw data files of ROPstat directly, without any conversion difficulty.
- With ROP-R, one can perform several special statistical analyses – mainly in the area of clustering – that are currently unavailable in JASP, jamovi or other well-known menu-driven statistical softwares.

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

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

² see <https://www.jamovi.org/features.html>

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

1. ROP-R can be started by running the program ROP-R.exe under Windows system 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 a successful installation, ROP-R.exe will be located in folder „c:_vargha\ropstat”. From here, one can run ROP-R directly or make a shortcut on the tray for it.
2. Since 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 then the path of program Rcmd.exe is to be specified in 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, but the path c:\Program Files\R\R-4.1.3\bin\x64\Rcmd.exe also works. The latter can run 64 bit procedures faster³. Newer R versions (4.2.0 and above) may fail to run with ROP-R properly⁴.
3. In the next step, several packages of R have to be installed as follows.
 - a. R has a basic program called RGui.exe (in the same folder as Rcmd.exe).
 - b. After starting RGui, copy the lines below in a regular way to its console and press Enter. After a couple of minutes, these packages will properly be installed and you can exit RGui.

```
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 first running only the first line of the list above (install.packages("cluster", dependencies = TRUE)), accepting the offered options; following the successful installation of the package *cluster* copy the remaining lines (R commands) and run. If any problem occurs with installing the packages together, try to install them separately.

³ see <https://askubuntu.com/questions/54296/difference-between-the-i386-download-and-the-amd64>

⁴ In newer versions of ROP-R the authors will try to overcome this difficulty.

4. Following these steps, ROP-R can be started and data files can be read the same way as in ROPstat. The default is ROPstat's msw file format. Apart from this, ROP-R accepts Excel type xls, xlsx files⁵, 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.
6. During running a module, ROP-R always runs one or more R scripts (with appropriate messages on the screen). After executing a script, the results are written into a text file. ROP-R automatically reads this and creates user-friendly tables for output that one can send or copy to Excel or Word.
5. An important feature of ROP-R is that the R-scripts run by ROP-R are always written into text files with *.r extension (such as EFA.r, CFA.r, PolReg.r, MBCA.r, etc.), available for 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 needed by the user.
6. If the results include figures, ROP-R places them into the folder mentioned above and informs about this fact in the end of the output.

2. The statistical menu system of ROP-R

In this section, we present the currently available statistical modules of 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, and work exactly the same way as in ROPstat (see *Vargha et al.*, 2015), and similarly to other softwares. For this reason, we do not deal with 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 may be sorted to separate blocks by means of block indices specified by the user. Block indices are integer numbers that may vary between 1 and 9. Variables having the same block index belong to the same block. The main purpose of HierR is to measure and test the effect a block adds to the set of preceding blocks in predicting a dependent variable.

The blocks enter the multiple linear regression model separately one after another, in consecutive steps. That is why this type of multiple regression is called hierarchical or sequential (*Tabachnick–Fidell*, 2013, p. 137). The final model is the simplest one after which

⁵ In reading Excel files in ROP-R the worksheet, containing the data table must be set to the active one.

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.
- R-squared value (R^2), the explained proportion of variance of the cumulative regression model.
- The adjusted R -squared value.
- The standard error of the regression estimation (RMSE) of the cumulative regression model.
- The significance test of the cumulative regression model (F statistic, df values, and p value).
- The increase of R^2 (R^2+) compared to the model not including 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 basic descriptive statistics of all specified variables, and displays, for each dependent variable, the regression coefficients of the final model (with their significances), 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 inflating factor. Multicollinearity is indicated by very high R^2 and VIF, and very low TOL values ($R^2 > 0.99$, $TOL < 0.01$, $VIF > 100$). Independent variables causing multicollinearity should be dropped from the analyses. The presence of multicollinearity may cause 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 the user whether the mediation analysis is to be performed. If the answer is yes, then beyond a standard HierR₂ each independent variable in the second block will be regarded – and analyzed – as a moderator variable in the relationship of 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 measuring the influential impact. We consider a data point highly influential if removing the point would substantially impact the regression model. In linear regression, Cook's distance (D) is often suggested to be calculated for each observation, in order to describe that observation's degree of influence upon the model (Marzjarani, 2015). Observations with larger D values have greater influence. Saving D , a robust regression analysis can be performed by omitting highly influential cases by regarding outliers 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 $D_{min} = 4/n$ (n = total number of cases) is a potential outlier, though this D_{min} threshold is not a rigorous standard. Modifying the default value (=1) of the "Multiplier of standard (4/n) threshold" option in the HierR module one can either increase or decrease the number of outliers. Checking in also the "Save influential case variable" option, the D/D_{min} 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 in folder `c:_vargha\ropstat\aktualis` the data file of independent and dependent variables (`tmpdat.txt`),

the R-script used (HierReg.r), the created path plots (or plots) of the mediation analysis (modplot*.pdf) and the text file of the R-output (oo.txt).

Polynomial regression (PolR)

In this module, each dependent variable is regressed on each independent variable. The main purpose of PolR is to find independent variables having nonlinear effect on some dependent variables. PolR is basically a HierR analysis where the powers of the independent variable ($X^1 = X, X^2, X^3, \dots$) 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 an evidence of a nonlinear effect. In a regression equation of PolR, X represents the linear component, X^2 the quadratic component, X^3 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 whose main feature is one single dense region of the multivariate data. Such a sample therefore may be appropriate for exploring several dense regions or value patterns with cluster analysis (Vargha *et al.*, 2016; Vargha–Bergman, 2019).

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

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. In BLR only dichotomous dependent variables are accepted, and each of them is regressed on the same set of specified independent variables. The independent variables can be either continuous (default) or categorical – 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 will be created for its different values⁶.

The blocks enter the regression model separately, one after another. The final model is the simplest one after which a newly entered block of independent (explanatory) variables does not increase significantly the fit of the regression model. BLR is basically a multiple linear regression where 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 dummifying), but it can also be regarded as a nonlinear discriminant analysis with a two-group categorical criterion variable (see *Tabachnick–Fidell*, 2013, Chapter 10). If p is a number between 0 and 1,

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

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

⁶ For each value and each case, the dummy variable is 1 if the case is characterized with this value and 0 if not.

- 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
- The significance test of the cumulative BLR model (χ^2 statistic, df value, p value)
- The significance test of the improvement of the given BLR model compared to the preceding one (χ^2 statistic, df value, p value).

The output of the model contains the basic descriptive statistics of all specified variables, and for each predictor the regression coefficients (and their significances) of the final model. This table also contains the odds ratio = $\exp(B)$ value of each X predictor, 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 odds – increases due to the unit increase of 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 (= value minimum) to the one indicated in the current row.

In addition, the output contains collinearity diagnostics (TOL and VIF measures) for the independent variables (independent variables causing multicollinearity should be dropped from the analysis), the classification table for the final model, and the summary table of the predictive measures (accuracy, specificity, and sensitivity percentages). The data of the latter two tables depend on the cut-off value of the probability prediction scale that 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`), the used R-script (`BinLogReg.r`), and the 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 with an orthogonal (Varimax) or an oblique (Promax) method. The output of PCA 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, that is the table of correlations between variables and rotated components (if oblique rotation is requested).

A special option of this module (based on the R package *MBESS*; see *Kelley*, 2007) is that one can request internal consistency reliability measures of the input variables. When this option is chosen, the program computes Cronbach's alpha and McDonald's omega reliability

measures for the input variables, regarding them as items of one single scale. It is important to know that if an item is to be reversed, this transformation must be performed on the item prior to the reliability analysis. Within 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 $c = \text{item minimum} + \text{item maximum}$), or using the item conversion option in the *Item analysis* module of ROPstat.

Another option in PCA is the listing of outlier cases and saving a variable measuring the case extremity, 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 center of the total sample, as a measure of case extremity, RMD can be saved as a new variable. By means of this variable, playing a similar role as 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 is 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 one can increase the number of outliers.

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

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) (see Osborn, 2014, *Tabachnick–Fidell*, 2013, Chapter 13). The extracted primary factor structure can be rotated with an orthogonal (Varimax) or an oblique (Promax) method. The output of 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, 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, TLI);
- Pattern matrix, that is, the matrix of rotated factor loadings (if rotation is requested);
- Correlation matrix of rotated factors (if oblique rotation is requested);
- Structure matrix, that 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 in folder `c:_vargha\ropstat\aktualis` the data file of the input variables (`tmpdat.txt`), and the used R-scripts (`Rbegin.r`, `EFA.r`).

Confirmatory factor analysis (CFA)

Based on R packages *lavaan* (Rosseel, 2012) and *lavaanPlot* (Lishinski, 2021), this module performs 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 to separate factors (scales) by means of factor indices specified by the user. These indices are integer numbers that may vary between 1 and 9. Variables having the same factor index belong to the same factor. The main purpose of CFA is to test the specified factor model (via χ^2 test) and measure the goodness of the model fit by means of 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;
- ULMSV: 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 a mean and variance adjusted test statistic.

CFA runs with the chosen estimator of the specified model and of 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 output of CFA contains the following standard results.

- Basic descriptive statistics;
- The modification indices of covariances of the specified model;
- χ^2 tests of the specified and the baseline model;
- Fit indices for the different methods and models (AIC, BIC, RMSEA, pClose, CFI, TLI, 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 standardized factor loadings of latent factors also 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 standardized 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 in folder `c:_vargha\ropstat\aktualis` the data file of the input variables (`tmpdat.txt`), the used R-scripts (`CFA.r` and `CFA2.r`), the text files of the R-output (`oo.txt` and `o2.txt`), and the pdf files of the prepared path diagrams (`pathplot1.pdf` and `pathplotR1.pdf` for the specified model, and the improved model respectively). 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 them unavailable in any 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 with 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 in one cluster (agglomerative clustering algorithm). On the contrary, 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 subclusters (divisive clustering algorithm).

In the nonhierarchical 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 multivariate normal. In this framework MBCA is the exploration of the underlying mixture structure, deciding on size (cluster number) and on the type of the cluster structure (*Fraley–Raftery*, 2002; *Gergely–Vargha*, 2021).

Despite all these possibilities, ROP-R does not cover all clustering methods available in ROPstat that 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 that is suggested to be used 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) by means of six alternative distance types (Squared Euclidean, Euclidean, Manhattan, Canberra, Maximum, Minkowski) and eight agglomerative methods (Average, Single, Complete, Centroid, Median, Ward, Flexible beta, McQuitty). Four optional diagrams (dendrogram, Silhouette plot, Total WSS plot, Banner diagram) may help the evaluation of the results.

The user may chose 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 may be saved (attached to the end of the actual msw data file).

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

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) by means of six alternative distance types (Squared Euclidean, Euclidean, Manhattan, Canberra, Maximum, Minkowski). Optional diagrams, the output, and saving options are the same as in AHCA.

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

k-center cluster analysis (KCA)

Based on R packages *stats* (R Core Team, 2021), *cluster* (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*-centers cluster analysis: *k*-means, *k*-medoids, and *k*-medians. The latter two methods are suggested when the input variables are seriously nonnormal or ordinal (Kaufman–Rousseeuw, 2009, Chapter 2). The best-known *k*-means analysis can be performed with three optional algorithms (Hartigan–Wong, MacQueen, Lloyd/Forgy). Several types of plots (Silhouette, EESS%, Mean heterogeneity, $f(K)$ distortion) help to determine the optimal cluster number. The structure of the output and the saving options are similar to those in AHCA.

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

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 model-based cluster analysis. The program searches for each model – belonging to different cluster numbers (specifiable between 2 and 25) and to maximally 14 possible model types – the best fit by means of an ML algorithm (Fraley–Raftery, 2002). The best model is the one having the largest BIC or ICL criterion value.

The evaluation of the results is based on the BIC or ICL plot, and the summary tables of the optimal solution. The structure of the output and the saving options are similar to those in KCA. Optional 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 in folder c:_vargha\ropstat\aktualis the data file of the input variables (tmpdat.txt), the data file extended with the cluster code variable for the optimal MBCA solution (tmpdat2.txt), the used R-script (MBCA.r), and the requested diagrams (e. g., BIC plot) in jpg files.

3. An illustrative example using ROP-R

In a study of psychological attachment, we obtained data on ECR-RS (Experiences in Close Relationships – Relationship Structures) questionnaire (Fraley *et al.*, 2011) from 336 Hungarian adult subjects (124 males and 212 females) living in a heterosexual relationship⁷ (Jantek–Vargha, 2016). This questionnaire assesses individual differences in attachment within and across a variety of relational contexts, among them the attachment toward mother, father, romantic partner, and friend. In the 40-item self-report questionnaire 10 items (6 measuring avoidance, 4 measuring anxiety) belong to each of the four domains with the same questions. In each domain, subscales Anxiety and Avoidance are formed from the corresponding items. In the model of attachment, anxiety and avoidance are the two main components. Persons having no attachment problems have a low level in both anxiety and avoidance, and the attachment is highly problematic if both anxiety and avoidance show a high level (Fraley *et al.*, 2011).

In what follows we present some steps of the Hungarian adaptation of ECR-RS, using several modules of ROP-R. For illustration purposes we chose the Partner scale with its 10 items: Av_P01, Av_P02, Av_P03, Av_P04, Av_P05, Av_P06, An_P07, An_P08, An_P09, An_P10 (here Av stands for avoidance, Anx for anxiety, P for partner).

1. First we performed a PCA on the above 10 items with the PCA module of ROP-R. We checked the reliability measures options and obtained the following results:

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

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

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

2. Next, we performed an EFA on the above 10 items with the EFA module of ROP-R. Since the 7-point items were all strongly nonnormal (either skewness or kurtosis or both were significant at $p < .001$ 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 can be seen in Table 1.

Table 1

Sorted rotated factor loading matrix of the 10 items of the Partner scale (loadings under .2 are not presented)

Item	Factor1	Factor2	Factor3
Av_P02	0.898		
Av_P03	0.832		

⁷ Much thanks to Gyöngyvér Jantek who agreed to use her data sample in the statistical analyses presented in the present paper.

Av_P01	0.765		
Av_P04	0.601		
Anx_P09		0.913	
Anx_P07		0.831	
Anx_P08		0.690	
Anx_P10		0.575	
Av_P05			1.003
Av_P06			0.591

Table 1 shows that Avoidance and Anxiety items form clearly different factors. An astonishing result is that two items (Av_P05 and Av_P06) of the Avoidance subscale form a different factor, independently from the other Avoidance items. This happened probably because these two items were formulated in a negative form (for this reason they had been converted prior to PCA and EFA). For this reason, we omitted items Av_P05 and Av_P06 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 8 positive items of the Partner scale had an excellent KMO value (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 (loadings under .2 are not presented)

Item	Factor1	Factor2
Av_P02	0.932	
Av_P03	0.819	
Av_P01	0.722	
Av_P04	0.630	
Anx_P09		0.928
Anx_P07		0.846
Anx_P08		0.648
Anx_P10		0.588

3. Following this we 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 a 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 ($p > .35$, see the last column of Table 3). This model had excellent fit indices (see the last column of Table 4), because the χ^2/df value (parsimony fit index) was well less than 3.5, RMSEA and RSMR less than .05, pClose $> .50$, and CFI, TLI $> .95$ (Hu–Bentler, 1999; Browne–Cudeck, 1993; Kline, 2005).

Table 3

Testing model fit (the improved models include covariances based on large modification indices)

Model:	Standard ML	Robust MLMV	Improved ML	Improved MLMV
χ^2 statistic:	65.58	40.35	26.9	18.5
df:	19	19	17	17
p value:	< .001	.003	.060	.358

Table 4

Model fit indices (the improved models include covariances based on large modification indices)

Model:	Standard ML	Robust MLMV	Improved ML	Improved MLMV
χ^2/df	3.45	2.12	1.58	1.09
RMSEA	.086	.058	.042	.016
CI_95 (RMSEA)	(.064; .110)	(.033; .084)	(.000; .071)	(.000; .054)
pClose	.005	.264	.638	.923
CFI	.966	.966	.993	.998
TLI	.950	.950	.988	.996
SRMR	.044	.044	.029	.029

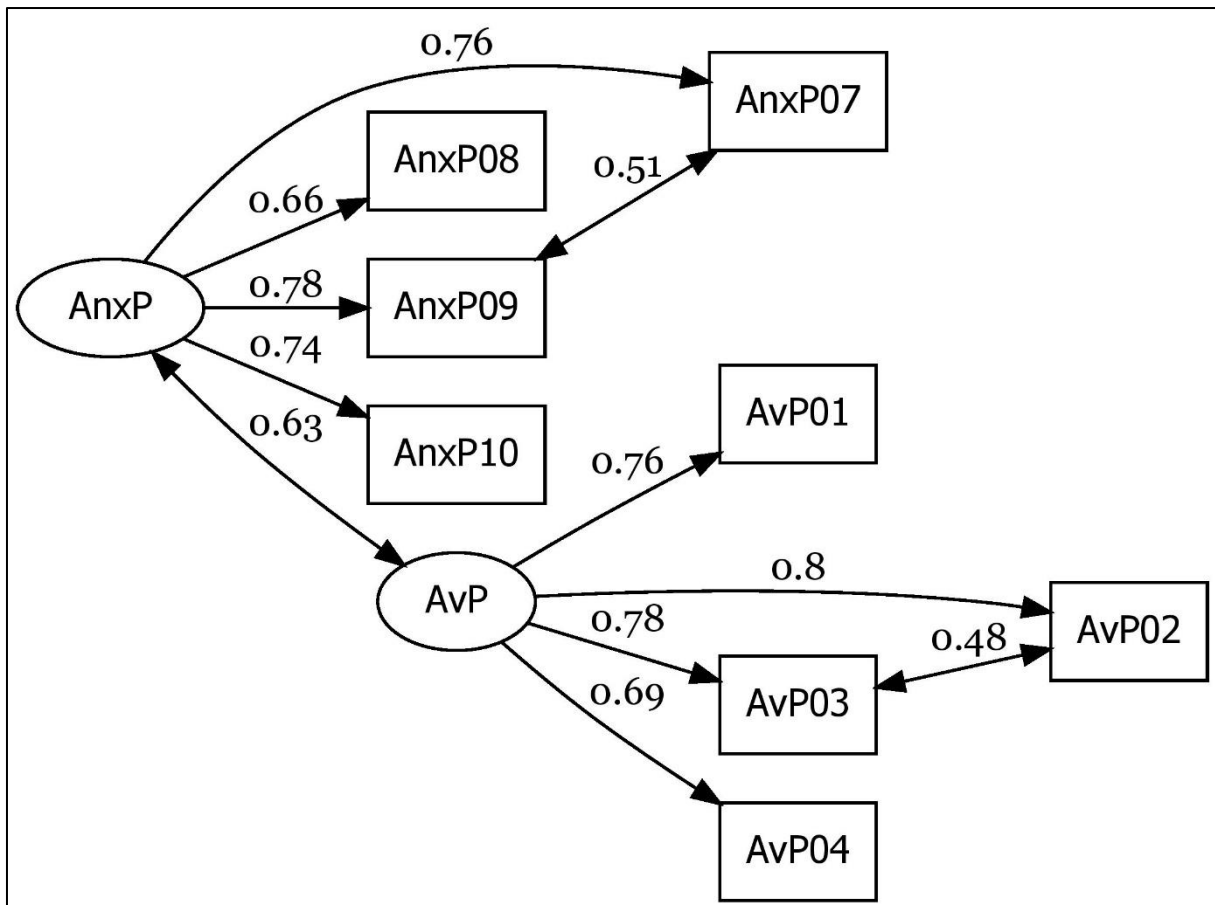


Figure 1: The path diagram with the standardized regression estimates of the Improved MLMV model of the Partner scale with two subscales: Anxiety (AnxP) and Avoidance (AvP).

The path diagram (with the standardized regression estimates) of the Improved MLMV model of the Partner scale with the two subscales: Anxiety (AnxP) and Avoidance (AvP) can be seen on Figure 1. ROP-R created a pdf file (with name pathplotR1.pdf) and this could be converted to a jpg file freely by means of website <https://pdf2jpg.net/>.

All the obtained results confirm that the Hungarian adaptation of ECR-RS has a good structural validity. An additional evidence for simplifying the model of the Partner scale of ECR-RS by omitting items Av_P05 and Av_P06 was that performing a CFA with all 10 original Partner items, Av_P05 and Av_P06 had the smallest standardized loadings (less than .30) in the improved MLMV model, whereas all the other standardized loadings were greater than .45. Having these good results, we created the Anxiety_P and the Avoidance_P subscales of the Partner scale by averaging the appropriate items, using a special option of module CFA, for the sake of performing further statistical analyses with these subscales.

4. Beyond structural reliability, it is important to prove the validity of an adapted scale. Several studies show that adult attachment and mental health are positively related. Individuals having a disordered attachment style (with high levels of anxiety and avoidance) usually have lower levels of mental health (see, e.g., *Adams et al.*, 2018). This was confirmed in our study as well, where the level of mental health was assessed by means of the 5-item WHO Well-Being Scale (WBI-5; see *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 .122 ($F(2; 321) = 22.28, p < .001$).

However, the question arises whether this influence of attachment still remains if basic personality traits of the Big Five model enter first the regression model. In our study the Big Five model was measured by means of the five scales (Extraversion, Agreeableness, Neuroticism, Conscientiousness, Openness) of the BFI-44 Big Five Inventory (*John et al.*, 2008). Performing a hierarchical regression analysis by means of module HierR of ROP-R, the first block was defined with the five BFI-44 scales, and the second block with the two attachment scales (Anxiety_P and Avoidance_P). The obtained results showed that the first block with the five BFI-44 scales explained a respectable 27.9% of the variance of WBI-5, but Anxiety and Avoidance, entering in a second block could increase this value by a significant ($F(2; 316) = 11.21, p < .001$) and explainable amount of .048 to a level of $R^2 = .327 = 32.7\%$.

5. Having reliable and meaningful Partner attachment subscales we want to explore basic attachment types using them as input variables in cluster analyses. If the input variables have a multivariate normal distribution that has only one single center then 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 contradict to multivariate normality.

Using Anxiety_P as dependent variable and Avoidance_P as 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 gives a 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	.258					
Quadratic	.268	.010	4.401	1	327	.037
Cubic	.280	.012	5.493	1	326	.020
4 th grade	.285	.005	2.218	1	325	.137
5 th grade	.306	.021	9.733	1	324	.002

6. Among cluster analyses, we first performed a standard Ward type agglomerative hierarchical analysis by means of the AHCA module of ROP-R, standardizing the input variables. We asked for dendrogram (see Figure 2), Silhouette plot (see Figure 3), and detailed results for cluster numbers 3 to 6. From these results some adequacy measures are summarized in Table 5. For each cluster the Homogeneity coefficient (HC) is the mean within cluster distance with the average squared Euclidean distance. The smaller HC is, the more homogeneous the cluster. HCmean is the average of the HC values (weighted with the cluster sizes) in a cluster solution (Vargha *et al.*, 2016). HCmin and HCmax denote the minimum and the maximum of the HC values. EESS%, the explained error sum of squares percentage is also a cohesion measure of a cluster structure. EESS% values above 65 indicate good structures. XBmod, the modified Xie-Beni index – similarly 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; Vargha *et al.*, 2017).

In AHCA the dendrogram (see Figure 2) does not indicate an 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). Results summarized in Table 6 show that EESS% has an acceptable level only above $k = 3$, and already for the $k = 4$ solution HCmean and XBmod have also good values. Only HCmax is too high, indicating that the cluster having the largest HC value is highly heterogeneous. Going with k a little upward (until $k = 6$), HCmax never decreases. For $k = 5$ the surprisingly low value of XBmod (= 0.094) may be due to two highly similar clusters in this solution, whose fusion in the $k = 4$ solution leads already to a good level of cluster separability for both XBmod (see Table 6) and the Silhouette index (see Figure 3).

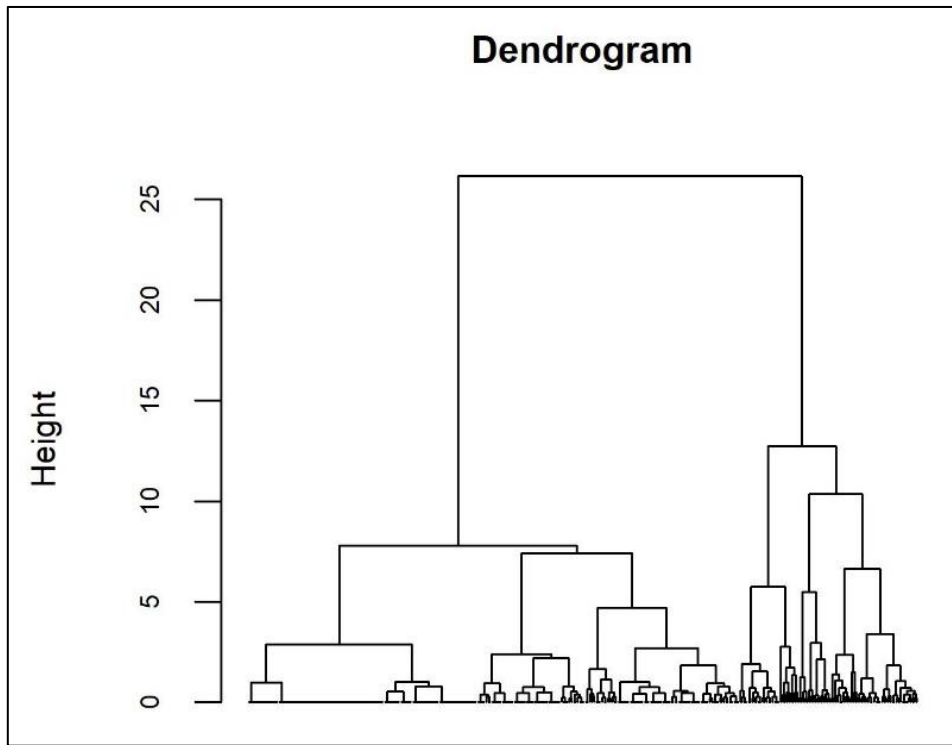


Figure 2: The dendrogram of Ward-type AHCA analysis with input variables Anxiety_P and Avoidance_P

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

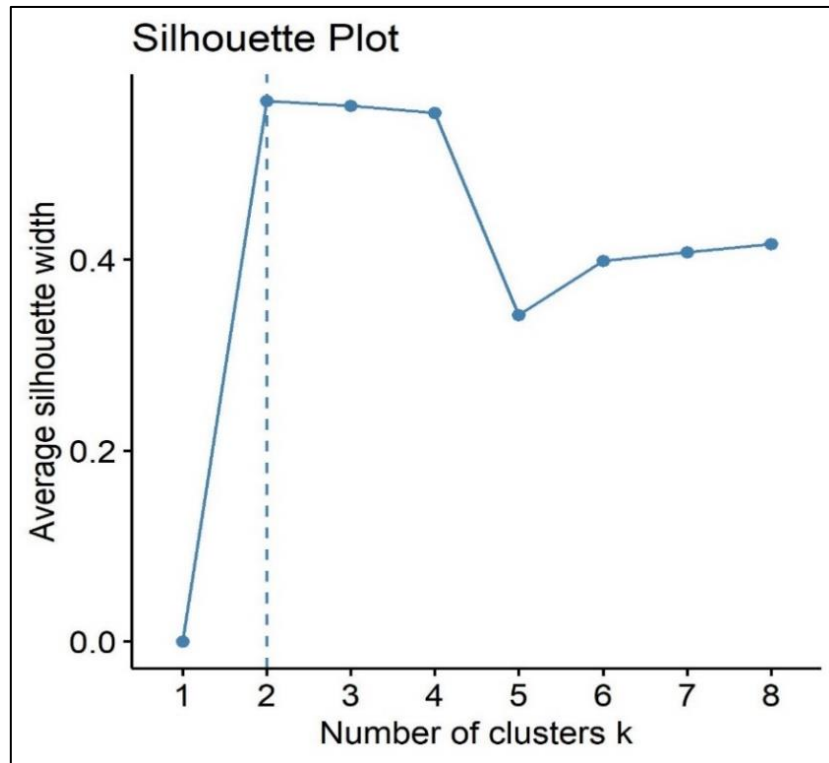


Figure 3: The Silhouette plot of Ward-type AHCA analysis with Anxiety_P and Avoidance_P

Based on the above detailed AHCA results the $k = 4$ solution seems to be the most promising. On the ROP-R output we find the table of standardized means of the explored clusters (see Table 7) that helps explaining the obtained cluster solution. Cluster CL1 is the largest and 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 the very high avoidance, the peculiarity of CL3 is the very high anxiety, and CL4 represents the most problematic attachment type with very high levels of both anxiety and avoidance. Though 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 (Fraleley *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. The model-based cluster analysis (performed with the MBCA module of ROP-R) did not yield a clear and well explainable solution either.

Performing k -center 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, created by means of the “Plot obtained clusters” option of the KCA module of ROP-R). On Figure 4 the four clusters are placed in the two-dimensional space of the first two principal components of Anxiety_P and Avoidance_P. On this figure cluster 4 matches well the highly homogeneous CL1 cluster in the four-cluster AHCA solution, cluster 3 the highly heterogeneous CL4 cluster, and clusters 1 and 2 the remaining two clusters of the four-cluster AHCA solution.

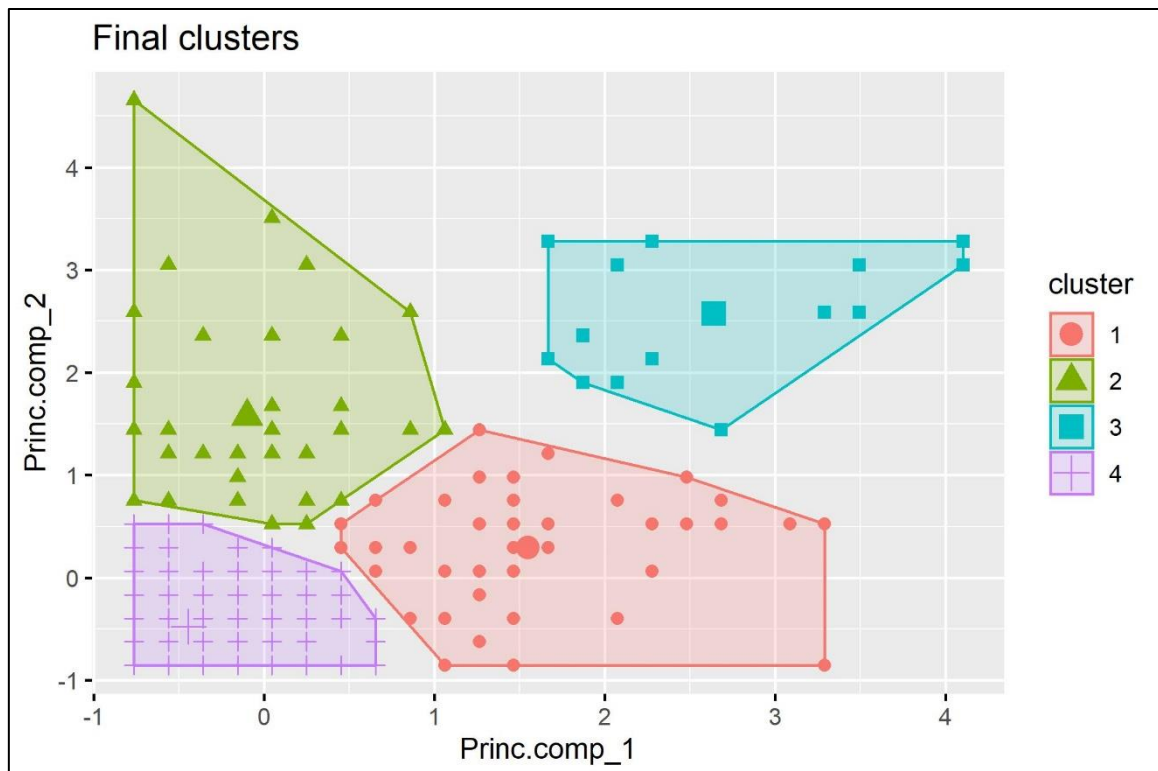


Figure 4: The four cluster solution of the k -means analysis with Anxiety_P and Avoidance_P

Discussion

In the present paper we introduced ROP-R, a free ROPstat based multivariate statistical program package that can be run under 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, just as some other recently developed softwares like JASP or jamovi. However, the modules of ROP-R offer several new and comfortable opportunities not available in other softwares. The usefulness and elegance of ROP-R was illustrated by a complex series of statistical analyses made on a sample of a study in the psychological topic of attachment.

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

- Install ROP-R from website <http://www.bansagi.hu/r/index.html>. After a successful installation ROP-R.exe will be found in folder c:_vargha\ropstat.

- Install R version R-4.1.3 for Windows from website <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 run with ROP-R properly.
- Start ROP-R.exe and specify the path of program Rcmd.exe in the menu point Settings/R-path (standard path is c:\Program Files\R\R-4.1.3\bin\x64\Rcmd.exe).
- Run RGui.exe (if there is not a shortcut on the desktop, 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 in point 3.b of Section 1 in the beginning of the present 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 get a positive feedback of the installation of package *cluster*, copy the remaining lines (R commands) from point 3.b of Section 1 in the beginning of the present paper and press Enter.

The above steps have to be performed only once. After completing them, you can enjoy running simply several modules of 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): *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. <http://www.stat.ucla.edu/research/gpa>
- 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*, 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. *Psychological 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 párok kapcsolatban élő felnőtt személyeknél. *Magyar Pszichológiai Szemle*, Vol. 71. No. 3. pp. 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. – 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* (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 (2nd ed.)*. New York, NY: Guilford Press.
- Lays, 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. 86-181). 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> Version = 2.2.5.
- 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]. Retrieved from <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]. Retrieved from <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): *Hungarian Statistical Review*. Vol. 2. No. 1. pp. 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*. Vol. 2. Nos. 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. 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>