

The
Statistical Menu
System
of

ROPstat

András Vargha
2007

Contents

1. Basics: single sample analyses

- Basic descriptive statistics
- Detailed statistics
- Frequency, histogram
- Tests on the mean and the median
- Test of normality

2. Comparison of groups and variables

- One-way independent samples ANOVA
- One-way ANOVA with repeated measures
- Two-way ANOVA
- Two-way mixed ANOVA
- Two-way rank ANOVA

3. Relationship of variables

- Correlation, simple regression
- Creation of correlation and partial correlation matrices
- Relationship of discrete variables
- Item analysis
- Multiple linear regression

4. Pattern-oriented analyses

- Description
- Imputation of missing values (Imputation)
- Residual analysis (Residue)
- Hierarchical cluster analysis (Hierarchical)
- Relocation: nonhierarchical K-means cluster analysis (K-means)
- Configuration frequency analysis (CFA)
- Cell-wise analysis of contingency tables (ExaCon)
- Centroid
- Dense point analysis (DensePoint)
- Time separation
- Time fusion

1. Basics: single sample analyses

Basic descriptive statistics

This module computes basic statistics for the selected variables, such as mean, SD, coefficient of variation (SD/mean), smallest and largest value (x_{min} and x_{max}), and smallest and largest standardized value (z_{min} and z_{max}).

Detailed statistics

This module computes some special statistics for the selected variables, such as median, mean, standard error of mean, confidence limits for the mean, skewness, kurtosis, and test of normality via skewness and kurtosis. If you set a nonzero trimming percentage (between 1 and 25, a more detailed output will be provided: mean, SD, trimmed mean, Winsorized SD, median; confidence limits for the theoretical mean, confidence limits for the theoretical trimmed mean, skewness, kurtosis, and test of normality via skewness and kurtosis.

Reference: Wilcox (1996, Chapter 2).

Frequency, histogram

This module computes univariate histogram and frequency distribution. The form of the output depends on the scale type of the variable.

- In the case of continuous scales the frequency distribution will be given to appropriately chosen classes instead of all particular values, unless the number of different values is small. Computed are frequencies, relative frequencies, and cumulative frequencies. The number of classes can be increased via increasing the class number multiplier constant. A histogram is made for this scale type.

- In the case of discrete scales each particular value is regarded as a distinct category. Computed are frequencies, relative frequencies, and cumulative frequencies. A histogram is also made for this scale type. For discrete variables it is possible to create and save special variables by means of each selected X variable.

(i) Binarization of X : for each found value c of variable X a new binary variable, X_c is created with $X_c = 1$ if $X = c$, and $X_c = 0$ if $X \neq c$.

(ii) Percentile transform of X : to each value c of variable X the sample cumulative frequency is assigned, that is the proportion of values less than or equal to c .

After the analysis these new variables appear in the last columns of the data matrix.

- If the Scale type is set to "group.def.", then information about the defined groups and their frequencies will be provided. Optionally the equality of theoretical frequencies can be tested.

Tests on the mean and the median

In this module for each selected variable the following tests are performed:

- one-sample t test and two robust alternatives (Johnson test and Gayen test);
- trimmed t -test (optional)
- Wilcoxon's signed rank test, and sign test.

The tested null hypothesis: the population mean (t -test and robust alternatives), the trimmed mean (trimmed t -test), or median (Wilcoxon and sign tests) equals the hypothetical value A specified in the program menu. For discrete variables the null hypothesis of the sign

test: $P(X > A) = P(X < A)$. Johnson and Gayen tests are performed only for sample sizes not increasing 500, and for $|t| \leq 10$.

Reference: Wilcox (1996, Chapters 7 and 15).

Test of normality

Normality test. This module performs a test of normality for each selected variable. In case of samples not greater than 100 the Kolmogorov-test, otherwise the chi-square test is performed.

In this module it is possible to create and save a standardized form of each selected variable.

(i) Simple standardization: in this case the mean and the SD of the new variable will be 0, and 1, respectively.

(ii) T-scale standardization: in this case the mean and the SD of the new variable will be 50, and 10, respectively.

After the analysis these new variables appear in the last columns of the data matrix.

2. Comparison of groups and variables

One-way independent samples ANOVA

In this module groups of independent samples can be compared. The statistical procedures applied depend on the type of scale of the dependent variables. For interval scales the module compares means and SD's via one-way ANOVA and its robust alternatives (Welch, James, and Brown-Forsythe methods). For ordinal scales stochastic homogeneity is tested. For nominal scales cross-tabulation is made and a chi-square test is performed. For performing a trimmed ANOVA (in the case of interval scales) set the trimming % to a value different from 0.

Reference: Maxwell & Delaney (2004, Chapter 3); Vargha & Delaney (1998); Wilcox (1996, Chapter 9).

One-way ANOVA with repeated measures

In this module variables (repeated measures) can be compared. For variables of interval scale the equality of means, trimmed means, as well as stochastic homogeneity are tested. For ordinal variables only stochastic homogeneity is tested. In the case of nominal variables the comparison of the distributions is carried out by discrete analyses (McNemar test, Cochran's Q-test), depending on the number of the dependent variables, and the number of found distinct values. For performing a trimmed ANOVA set the trimming % to a value different from 0.

Reference: Maxwell & Delaney (2004, Chapter 11); Wilcox (1996, Chapter 11).

Two-way ANOVA

Two-way comparison of independent samples. This module makes comparisons (by means of two-way ANOVA) among group means categorized by two grouping variables. For performing a trimmed two-way ANOVA set the trimming % to a value different from 0.

Reference: Maxwell & Delaney (2004, Chapter 7); Wilcox (1996, Chapter 10); Wilcox (2003, Chapter 10).

Two-way mixed ANOVA

This module performs a two-way analysis of variance with one grouping factor (represented by a grouping variable) and one trial factor (represented by a set of dependent variables).

Reference: Maxwell & Delaney (2004, Chapter 12); Wilcox (2003, Chapter 11).

Two-way rank ANOVA

This module makes stochastic comparisons, performing two-way rank ANOVA, between independent groups categorized by two grouping variables. With the main effects of this rank ANOVA the one-way stochastic homogeneity of both grouping factors can be tested. The stochastic interaction between the two factors means that the pattern of stochastic dominances of one of the factors is not the same at different levels of the other factor.

Reference: Brunner & Puri (2001), Kulle (1999).

3. Relationship of variables

Correlation, simple regression

This module computes simple linear regression and several types of pairwise correlations. If only X or only Y variables are selected, the analysis will be performed for all possible pairs within the set of the selected variables. If both X and Y variables are specified, then only the relationship between the X and Y variables will be analyzed (within X or Y variables not). Using the 'Cross' option each X variable will be related to each Y, while using the 'Pair' option only corresponding variables will be related (provided that the number of X and Y variables are equal).

Reference: Wilcox (1996, Chapter 13).

Creation of correlation and partial correlation matrices

This module prepares matrices of correlations and partial correlations of different types. If only X or only Y variables are selected, correlations for all possible pairs within the set of the selected variables. If both X and Y variables are specified, then only the correlations between the X and Y variables will be analyzed (within X or Y variables not). Using the 'Cross' option each X variable will be related to each Y, while using the 'Pair' option only corresponding variables will be related (provided that the number of X and Y variables are equal).

Reference: Wilcox (1996, Chapter 13).

Relationship of discrete variables

In this module two-dimensional frequency tables are created, independence of the two variable is tested (chi-square test), and strength of their relationship is measured (Cramér's V contingency coefficient). Frequency tables are formed based on the groups assigned to the variables. If group definition is missing, default group creation is being made (each distinct value defines a separate group).

Reference: Wilcox (1996, Chapter 14).

Item analysis

This module performs a traditional reliability analysis of a single scale. During this the scale's internal consistency is measured using Cronbach's alpha as well as the adequacy of each item is assessed by item-total and item-remainder correlations (item analysis).

Reference: Allen, & Yen (2002).

Multiple linear regression

This module performs a multiple linear regression and computes multiple linear correlation (R).

Reference: Wilcox (1996, Chapter 13).

4. Pattern-oriented analyses

Description

This module gives descriptive information of a data set focusing on the configurations of missing values. The output includes basic descriptive statistics, pairwise correlations, and reports about dropouts for single variables and pairs of variables.

Imputation of missing values (Imputation)

This module performs imputation of missing values by means of three different methods: 1. Replacement with the variable mean. 2. Replacement with the most similar complete case (twin/nearest neighbor); 3. Replacement with the multiple linear regression estimation based on the selected variables.

Reference: Bergman, Magnusson and El-Khoury (2002, pp. 107-111).

Residual analysis (Residue)

This module identifies residual cases (outliers) based on the distance from the most similar case (cases), called twins, and creates a residual indicator variable by means of which these cases can be omitted from different classification analyses.

Reference: Bergman, Magnusson and El-Khoury (2002, pp. 109-110).

Hierarchical cluster analysis (Hierarchical)

This module performs an agglomerative hierarchical cluster analysis with optional relocation (k-means clustering).

Reference: Bergman, Magnusson and El-Khoury (2002, Chapter 4, and pp. 113-115).

Relocation: nonhierarchical K-means cluster analysis (K-means)

This module improves a cluster solution by the relocation of cases. It starts from an initial classification (it may also be a solution of a hierarchic clustering saved by module Hierarchical), and moves cases from one cluster to another if this leads to a reduction in the total error sum of squares of the cluster solution. In this way, bad-fitting cases are moved to a better fitting cluster and more homogeneous clusters can be obtained.

Reference: Bergman, Magnusson and El-Khoury (2002, Chapter 4, and pp. 115-117).

Configuration frequency analysis (CFA)

CFA: In this module configural frequency analysis is performed for identifying value patterns of a small set of discrete variables using the exact binomial test. Tests are made for both types and antitypes.

Reference: Bergman, Magnusson and El-Khoury (2002, Chapter 5, and pp. 117-119).

Cell-wise analysis of contingency tables (ExaCon)

In this routine exact cell-wise analysis is made of two-way frequency tables. Contingency table for all pairs of specified row and column categorical variables are created and analyzed with a focus on cell-wise types or antitypes based on exact tests (preferred are usually the two-tailed hypergeometric probabilities).

Reference: Bergman, Magnusson and El-Khoury (2002, Chapter 5, and pp. 125-127).

Centroid

Centroid comparison of different groups and cluster solutions. This module can compare clustering solutions with each other by matching each cluster centroid in one solution to the cluster centroid which resembles it the most in the other solution. The outcome is a set of pairs of centroids each belonging to one solution where the pairs are given in order of decreasing similarity. This module can also compare clustering solutions represented by their centroids (mean patterns) belonging to different groups.

Reference: Bergman, Magnusson and El-Khoury (2002, Chapter 5, and pp. 125-126).

Dense point analysis (DensePoint)

This module searches for objects with many neighbors. A dense point is defined as a case having many neighbors. If a control clustering variable is specified, the program reports the percentages of dense point neighborhoods falling into the different clusters for each explored dense point. In this case use a greater number (50-100) for the maximal number of dense points.

Time separation

Creates file where data at different time points are treated as subindividuals.

Time fusion

Recreation of complete individuals from a time separated file

References

- Allen, M. J. & Yen, W. M. (2002). *Introduction to Measurement Theory*. Long Grove, IL: Waveland Press.
- Bergman, L. R., Magnusson, D., & El-Khoury, B. M. (2002). *Studying individual development in an interindividual context. A Person-oriented approach*. Mahwah, New Jersey, London: Lawrence-Erlbaum Associates.
- Brunner, E. & Puri, M. L. (2001). Nonparametric methods in factorial designs. *Statistical Papers*, 42, 1-52.
- Kulle, B. (1999). *Nichtparametrisches Behrens-Fisher-Problem im Mehr-Stichprobenfall*. Diplomarbeit. Institut für Mathematische Stochastik der Georg-August-Universität Göttingen.
- Maxwell, S. E. & Delaney, H. D. (2004). *Designing experiments and analysing data. A model comparison perspective (2nd ed.)*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Vargha, A., & Delaney, H. D. (1998). The Kruskal-Wallis test and stochastic homogeneity. *Journal of Educational and Behavioral Statistics*, 23, 170-192.
- Wilcox, R. R. (1996). *Statistics for the social sciences*. San Diego, New York: Academic Press.

Wilcox, R. R. (2003). *Applying contemporary statistical techniques*. San Diego, New York: Academic Press.