R-functions for the analysis of variance

The following R functions may be downloaded from the directory
http://www.uni-koeln.de/~luepsen/R/

All file names of the functions have the extension ...`R`.

A complete set of all functions contains the file `anova.lib`, which can be made available after a download by means of `attach(`*`path`*`/anova.lib)`.

Usage advices:

- Variables used as factors have to declared as „factor“ by the user.
- The following command should be entered before usage:
  `options (contrasts=c("contr.sum","contr.poly"))`
- In contrary to most other R anova functions all variables must be part of the dataframe.
- data frames with missing values are not allowed, neither for repeated measures nor for between subject analyses.
  In such cases the function `na.omit` should be applied before performing the analysis.

There is warranty for correct functions.

Haiko Lüpsen, (last revision 12.5.2021)

Index

1. box.f: Box-F-Test for heterogeneous variances

1- or 2-factorial analysis of variance using the robust Box F-tests for heterogeneous variances.

Call: `box.f (model, dataframe)`

Parameter:

model	anova model (as in function aov) example: x ~ A*B
dataframe	Data, object of type data.frame

Result:

anova table	object of type data.frame and anova

Method:

Let there be *k* groups (main or interaction effect) with variances s_i^2. The F-test

$$F = \frac{MS_{Effekt}}{MS_{Fehler}}$$

can be corrected according to the heterogeneity of the variances by adjusting (exactly reducing) the numerator and denominator degrees of freedom. The numetrator df *df1* are multiplied with ε_1 , the denominator df *df2* with ε_2. The corrections are computed as followst:

$$\bar{s}^2 = \left(\sum s_i^2\right)/k$$

$$c^2 = \left(\sum (s_i^2 - \bar{s}^2)^2\right)/(k \cdot \bar{s}^4)$$

$$\varepsilon_1 = \left(1 + \frac{k-2}{k-1}c^2\right)^{-1} \qquad \varepsilon_2 = (1 + c^2)^{-1}$$

Here $\bar{s}^2$ may be interpreted as an average variance and c^2 as a dispersion of thevariances. It is easy to recognize that in the case of equal variances c^2=0 results and therefore ε_1 and ε_2 become 1.

References:

B.J. Winer et.al. (1991): *Statistical Principles in Experimental Design,* p 109

2. bf.f: Brown & Forsythe-F-test for heterogeneous variances

1- or 2-factorial analysis of variance using the robust method of Brown & Forsythe for heterogeneous variances.

The function consists of bf.f, bf2.f, bf3.f, bf.main, bf.orthog, bf.fratio.

Call: `bf.f (model, dataframe, mod=T)`

Parameter:

model	anova model (as in function aov) example: x ~ A*B
dataframe	Data, object of type data.frame
mod	T: modification of error degrees of freedom by Mehrotra is used F: original version of error degrees of freedom by Brown & Forsythe

Result:

anova table	object of type data.frame and anova

Method:

1-factorial analysis:
Let there be *k* groups with variances s_i^2, cell frequencies n_i . Brown & Forsythe use the following ratio which follows approximately an F distribution:

$$F = \frac{SS_{Effekt}}{SS_{Fehler}}$$

Here SS_{Error} is computes as (with $n = \sum n_i$)

$$SS_{Error} = \sum\left(1 - \frac{n_i}{n}\right)s_i^2$$

The denominator degrees of freedom of the F-test are computed as follows

$$df = \left(\sum\frac{m_i^2}{n_i - 1}\right)^{-1} \qquad m_i = \left(1 - \frac{n_i}{n}\right)s_i^2/(SS_{Error})$$

The 2-factorial analysis requires much more computation. For details the reader is referred to the article cited below.

References:

Brown & Forsythe (1974): *The Anova and Multiple Comparisons for Data with Heterogeneous Variances*, Biometrics, Vol. 30, No. 4, pp. 719-724

Devan V. Mehrotra (1997): *Improving the Brown-Forsythe solution to the generalized Behrens-Fisher problem*, Communications in Statistics - Simulation and Computation, 26:3, pp. 1139-1145.

3. mbf.f: modified Brown & Forsythe-F-test for heterogeneous variances

2-factorial split-plot analysis of variance using the robust method of Brown & Forsythe for heterogeneous variances in the version of Vallejo et al., which allows nonsphericity and heterogeneous covariance matrices. `mbf.f` allows data in wide as well as in long format.

The function requires the function bf.f (from this package).

Call (wide format): `mbf.f (`*`df, group`*`)`

Call (long format): `with(df,mbf.f (`*`y, groups=g, trial=w, id=id`*`)`

Parameter:

df	Dataframe, object of type data.frame, in case of wide format: only the repeated measurement variables
group	grouping factor
trial	repeated measurement factor
id	case id

Result:

anova table	object of type data.frame and anova

Examples:

Call (wide format): `mbf.f(winer518[,3:5],winer518,2])`

Call (long format): `with(winer518t, mbf.f (`*`y, groups=Geschlecht, trial=Zeit, id=Vpn`*`)`

References:

Vallejo, G., Moris, J., & Conejo, N. M. (2006).
A SAS/IML program for implementing the modified Brown–Forsythe procedure in repeated measures designs.
computer methods and programs in biomedicine, 83(3), 169-177.

Guillermo Vallejo, P. Fernández, F. J. Herrero and N. M. Conejo (2004):
Alternative procedures for testing fixed effects in repeated measures designs when assumptions are violated.
Psicothema . Vol. 16, nº 3, pp. 498-508

4. wj.anova: Welch-James-anova for heterogeneous variances in between subject designs

1- or 2-factorial analysis of variance for between subject designs using the robust method of Welch & James for heterogeneous variances

Call: `wj.anova (dataframe, dependent variable, grouping factors, Ftest)`

Parameter:

dataframe	Data, object of type data.frame
dependent variable	variable name (in "...")
grouping factor 1	variable name (in "...")
grouping factor 2	variable name (in "..."), optional
Ftest	F (default): chi^2-test as proposed by Algina & Olejnik T: F-test as proposed by Lix, Algina & Keselman

Result:

anova table	object of type data.frame

Example:

```
wj.anova (mydata,“y“, “g1“, “g2“)
```

References:

James Algina & Stephen Olejnik (1984): *Implementing the Welch-James procedure with factorial designs*, Educational and Psychological Measurement, 44, 39-48

Lix, L. M., Algina, J., & Keselman, H. J. (2003). *Analysing multivariate repeated measures designs: A comparison of two approximate degrees of freedom procedures*. Multivariate Behavioral Research, 38, 403–431.

5. wj.spanova: Welch-James-anova for heterogeneous covariance matrices in split plot designs

1- or 2-factorial analysis of variance for mixed designs (split plot designs) using the robust method of Welch & James for heterogeneous covariance matrices. The dataframe must have the same structure as it is requested by `aov` oder `ezANOVA`.

Call: `wj.spanova (`*`dataframe, dep var, group factor, repmes factor id variable`*`)`

Parameter:

dataframe	Data, object of type data.frame
dependent variable	variable name (in "...")
grouping factor	variable name (in "...")
rep.meas. factor	variable name (in "...")
id variable	variable for identifying cases (in "...")

Result:

anova table	object of type data.frame

Example:

```
wj.spanova (mydata,"y", "group", "time", "Vpn")
```

References:

H.J.Keselman, K.C.Carriere und L.M.Lix: *Testing Repeated Measures Hypotheses When Covariance Matrices are Heterogeneous*, Journal of Educational Statistic, 1993, Vol. 44, No 4

6. box.andersen.f: F-test for nonnormal distributed variables

1- or 2-factorial analysis of variance using the robust method of Box & Andersen for nonnormal distributed dependent variables.

Call: `box.andersen.f (model, dataframe)`

Aufrufparameter:

model	anova model (as in function aov) example: x ~ A*B
dataframe	Data, object of type data.frame

Result:

anova	anova table: object of type data.frame and anova
eps	factor for the correction of the degrees of freedom

Method:

Box and Andersen use a modified F-test where teh numerator and denominator degrees of freedeom are multplied with a parameter d. This is computed by means of the variance and kurtosis of the variable x. The following computation of the correction d is valid for approximately equal n_i. Therefore let n be the number of observataions in each group. (It shopuld be remarked that there exists also a more extensive formula for strongly varying n_i.

.

$$S_2 = \sum_i^k \sum_i^n (x_{ij} - \bar{x})^2 \qquad S_4 = \sum_i^k \sum_i^n (x_{ij} - \bar{x})^4$$

These sums are used to compute the following intermediate parameters:

$$k_2 = S_2 / (n-1)$$

$$k_4 = [n(n+1)S_4 - 3(n-1)S_2^2] / [(n-1)(n-2)(n-3)]$$

And finally d is computed as

$$d = 1 + \frac{1}{n} \frac{k_4}{k_2^2}$$

References:

Dieter Rasch & Moti Lal Tiku: *Robustness of Statistical Methods and Nonparametric Statistics*, Reidel Publishing, 1984

7. check.covar: test of homogeneity of covariance matrices

Several tests of homogeneity of covariance matrices:

- Likelihood Ratio test,
- Box M-test, based on multivariate normal distribution,
- Schott‘s T1, an improvement of Box M-test, based on multivariate normal distribution,
- Schott‘s T2, for elliptic distributions, incorporating their kurtoses for measuring the deviation from the multivariate normal distribution, assuming all distributions als equal,
- Schott‘s T3, for elliptic distributions, incorporating their kurtoses, allowing different distributions,
- multivariate Levene test, assuming at least an ordinal scale (details see below),
- a pure dispersion test.

The T2 and T3 allow a wider area of distributions and are generalisations of T1. In his paper Schott recommends T2 for general usage, but limited to balanced designs. But own simulations showed that only the Levene test controls the type error rate in nearly all situations (e.g. different distributions) and has on the other side a good power performance in case of variance heterogeneity. For details see Luepsen (2020).

The function accepts the data in wide format as well as in long format.

Call (wide format): `check.covar (`*`dependent variables, grouping variable`*`)`
Call (long format): `check.covar (`*`dep varianle, groups=.., trial=.., id=..`*`)`

Parameters (wide format):

`dependent variables`
: repeated measurement variables as a dataframe or matrix

`grouping variable`
: vector containing the values of the grouping factor

Parameters (long format):

`dependent variable`
: vector containing the values of the repeated measurement variable

`grouping variable`
: vector containing the values of the grouping factor

`trial factor`
: vector containing the values of the trial factor

`case id`
: vector containing the case ids

Examples:
```
check.covar (winer[ ,c("V3","V4","V5")], winer$V2)
with(winer,check.covar(v,sex,time,id))
```

Output:

dataframe with χ^2-values, df and p value for the 6 tests, NAs if a test is not computable, e.g.

if there are too less cases in relation to the number of dependent variables.

References:

Hallin, Marc & Paindaveine, Davy (2009):
Optimal tests for homogeneity of covariance, scale, and shape.
Journal of Multivariate Analysis, 100, pp 422-444.

O‘Brien, Peter C. (1992): Robust Procedures for Testing Equality of Covariance Matrices.
Biometrics, Vol. 48, No. 3 (Sep., 1992), pp. 819-827

Luepsen, Haiko (2020):
Checking the Homogeneity of Covariance Matrices: some practical aspects.
URL: http://www.uni-koeln.de/~luepsen/statistik/texte/Checking.Homogeneity.pdf

Levene test:

For i=1,...,I groups, k=1,...,n_i , N ($N = \sum n_i$) subjects and j=1,..,J repeated measurements $y_{k(i)j}$ shall denote the j th measurement of variable y for subject $k(i)$ and m_{ij} the median in group i. Then for each subject $k(i)$ the following covariances $s_{j_1 j_2}$ are computed:

$$s_{k(i)j_1j_2} = (y_{k(i)j_1} - m_{ij_1})(y_{k(i)j_2} - m_{ij_2}) \quad (j_1, j_2 = 1,...,J)$$

which are transformed into

$$\hat{s}_{k(i)j_1j_2} = \mathrm{sgn}(s_{k(i)j_1j_2}) \cdot \sqrt{|s_{k(i)j_1j_2}|}$$

where sgn denotes the sign function. In the next step for each subject $k(i)$ the lower triangular of $\hat{s}_{k(i)j_1j_2}$ is transformed into a vector which results in a data matrix Y with N rows and $(J+1)J/2$ columns. Finally a multivariate analysis of variance, e.g. Wilks Lambda test, is applied on Y which checks the homogeneity of the covariance matrices of groups i=1,...,I.

8. check.corr: test of homogeneity of correlation matrices

Several tests of homogeneity of correlation matrices:

- Jennrich Test,
- Larntz & Perlman Test,
- a modified, restricted to correlations, Levene like test by O‘Brien (see above),
- a modified, restricted to correlations, Box-M-Test

Details will be found in Luepsen (2020), see check.covar.

The function accepts the data in wide format as well as in long format.

Call (wide format): `check.corr (`*`dependent variables, grouping variable`*`)`
Call (long format): `check.corr (`*`dep varianle, groups=.., trial=.., id=..`*`)`

Parameters (wide format):

`dependent variables`
repeated measurement variables as a dataframe or matrix

`grouping variable`
vector containing the values of the grouping factor

Parameters (long format):

`dependent variable`
vector containing the values of the repeated measurement variable

`grouping variable`
vector containing the values of the grouping factor

`trial factor`
vector containing the values of the trial factor

`case id`
vector containing the case ids

Examples:
```
check.corr (winer[ ,c("V3","V4","V5")], winer$V2)
with(winer,check.corr(v,sex,time,id))
```

References:

Box, G.E.P. (1949): A General Distribution Theory for a Class of Likelihood Criteria, Biometrika, Vol. 36, No. 3/4, pp. 317-346

Larntz, Kinley & Perlman, Michel D. (1985): A simple Test for the Equality of Correlation Matrices. University of Washington, Technical Report No. 63.

Jennrich, Robert I. (1970): An Asymptotic chi2 Test for the Equality of Two Correlation Matrices. Journal of the American Statistical Association, Vol. 65, No. 330, pp 904-912

O‘Brien, Peter C. (1992): Robust Procedures for Testing Equality of Covariance Matrices. *Biometrics*, Vol. 48, No. 3 (Sep., 1992), pp. 819-827

9. check.sphere: test for sphericity of a covariance matrix

Several tests for sphericity of a covariance matrix:

- Likelihood ratio test,
- a test by Muirhead & Waternaud (1980) for elliptic distributions, incorporating their kurtoses for measuring the deviation from the multivariate normal distribution, allowing a wider area of distributions (a generalization of the LR test),
- John‘s V test (see John, 1972), approximated by a beta distribution,
- John‘s V test, considering the deviation of the kurtosis from the value for the normal distribution (Hallin & Paindaveine, 2006), approximated by a χ^2-distzribution
- John‘s V test, using a more accurate computation of the p value (see Nagao, 1973),
- Mauchly‘s test (e.g. Winer, 1991, p. 255),
- multisample Mauchly test (Mendoza, 1980),
- multisample Mauchly test (Harris, 1984),
- a test for compound symmetry, as described by Winer (1991, p. 517).

The most favorable tests are said to be John‘s and Nagao‘s tests, though the two for elliptic distributions offer a wider range of applications. One remark to John‘s test: many different versions of this test can be found in the publications. Some of them differ only in the presentation of the formulae, but some also lead to different results. Therefore the formula used is presented below. For more information see Luepsen (2020).

The function accepts the data in wide format as well as in long format.

Call (wide format): `check.sphere (`*`dependent variables [, grouping variable]`*`)`
Call (long format): `check.sphere (`*`dep variable, trial=.., Id=.. [, groups=..]`*`)`

Parameters (wide format):

`dependent variables`
repeated measurement variables as a dataframe or matrix

`grouping variable`
vector containing the values of the grouping factor (optional)

Parameters (long format):

`dependent variables`
vector containing the values of the repeated measurement variable

`groups`
vector containing the values of the grouping factor (optional)

`trial`
vector containing the values of the trial factor

`Id`
vector containing the case ids

Example:
```
check.sphere (winer[ ,c("V3","V4","V5")])
with(winer,check.sphere(v,time,id))
```

Output:

`$results:` dataframe with χ^2-values, df and p value for the 8 tests.
For John's test with the kurtosis correction the column chisquare contains the kurtosis .

`$Box.epsilon:` the value of Box ε

`$k.deviation:` deviation of the kurtosis from the value of the normal distribution

`$error:` error code:
2: data matrix and vector of grouping variable have different lengths.

References:

Winer et al. (1991): *Statistical principles in experimental Design*, 1991.

Hallin, Marc & Paindaveine, Davy (2006): *Optimal Rank-Based Tests for Sphericity,* The Annals of Statistics, Vol. 34, No. 6, pp 2707–2756

Harris, P. (1984): *An Alternative Test For Multisample Sphericity.* Biometrika, No. 49, 2, pp 273-275.

John, S. (1972): *The distribution of a statistic used for testing sphericity of normal distributions*. Biometrika (1972), 59, 1, p. 169-173.

Li, Zeng & Yao, Jianfeng (2016): *Testing the sphericity of a covariance matrix when the dimension is much larger than the sample size*. Electronic Journal of Statistics, Vol. 10, pp 2973–3010

Luepsen, Haiko (2020): *Anmerkungen zum Testen der Spherizität*
URL: http://www.uni-koeln.de/~luepsen/statistik/texte/sphericity.pdf

Mendoza, Jorge L. (1980): *A significance test for multisample sphericity*, Psychometrika, Vol 45, No 4

Nagao, S. (1973): *On Some Test Criteria for Covariance Matrix.* The Annals of Statistics, Vol. 1, No. 4, pp. 700-709

Wang, Q. & Yao, J. (2013): *On the sphericity test with large-dimensional observations*. Electronic Journal of Statistics, Vol. 7

Muirhead, R.J. & Waternaud, C.M. (1980): *Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations*. Biometrika, 67

John's test:

Let S (s_{ij}) be the covariance matrix of the K repeated measurements. The statistic is computed in several steps:

$$U = \left(\sum_i^K \lambda_i^2\right) / \left(\sum_i^K \lambda_i\right)^2$$

where λ_i are the eigenvalues of the matrix CSC' with a normalized orthogonal transformation matrix C. Next the value

$$T = (K \cdot U - 1)/(K - 1)$$

is computed, for which several approximate distributions exist. Normally the χ^2-distribution is used, noting that this is only satisfying for larger samples *n*:

$$X = \frac{n}{2} \cdot K \cdot (K-1) \cdot T$$

X is χ^2-distributed with $(K\text{-}1)K/2$ - 1 degrees of freedom. For smaller samples also the statistic T can be used which is approximately beta-distributed or which can be transformed in an approximately F-distributed statistic. For details see John (1972).

As $U = 1/(K \cdot \varepsilon)$, where ε is the Box correction factor measuring the deviation from sphericity, above *X* can be written as

$$X = \frac{n}{2} \cdot K \cdot \left(\frac{1}{\varepsilon} - 1\right)$$

which directly shows that *X* is a statistic for the deviation from sphericity and e.g. X=0 for ε=1.

10. ats.2 : 2-factorial analysis of variance ats.3 : 3-factorial analysis of variance

2-factorial or 3-factorial analysis of variance using the ATS (anova type statistic) method of Akritas, Arnold and Brunner. Empty cells are not allowed.

Call: `ats.2 (`*`model, dataframe`*`)`
`ats.3 (`*`model, dataframe`*`)`

Parameter:

model	anova model (as in function aov) example: x ~ A*B
dataframe	Data, object of type data.frame

Result:

anova	anova table: object of type data.frame

References:

Edgar Brunner, Ullrich Munzel (2013): *Nichtparametrische Datenanalyse, Unverbundene Stichproben*, Springer, 126 ff.

Note:

For repeated measures designs there exists the package `nparLD` on Cran.

11. np.anova: nonparametric analysis of variance using the KWF-, Puri & Sen- and van der Waerden-method

factorial analysis of variance, with and without repeated measurements, either using the KWF-method, which is a generalisation of the well known Kruskal-Wallis- and Friedman-analyses, the Puri & Sen (L-statistic) method, or the van der Waerden method based on the inverse normal transformation. In the case of repeated measurements the dataframe must have the same structure as it is requested by `aov` or `ezANOVA`.

Call: `np.anova (`*`model, dataframe`*`)` generalized Kruskal-Wallis-Friedman
`np.anova (`*`model, dataframe, method=1`*`)` van der Waerden method

Parameter:

model	anova model (as in function aov) examples: x ~ A*B or score ~ group*time+Error(Vpn/time)
dataframe	Data, object of type data.frame (long format)
method	0: KWF generalized Kruskal-Wallis-Friedman tests 1: generalized van der Waerden method 2: Puri & Sen method 3: Puri & Sen method based on normal scores
compact	for repeated measurements only: T: all tests in one table (dataframe) (default) F: one table (dataframe) for each error term (similar to `summary(aov)`)
pseudo	ranking method: F: standard ranking (default) T: pseudo ranks

Result:

anova table	object of type data.frame and anova

References:

Puri, M.L. & Sen, P.K. (1985): *Nonparametric Methods in General Linear Models*, Wiley, New York

H. Mansouri and G. H. Chang (1995): *A Comparative Study of Some Rank Tests for Interaction*, Computational Statistics and Data Analysis, 19, 85-96

David J. Sheskin (2004): *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman & Hall

Lüpsen, Haiko (2020): *Some rank based ANOVA procedures for analyzing data from split-plot designs*. URL: http://www.uni-koeln.de/~luepsen/statistik/texte/algorithm.pdf.

Brunner, E., Konietschke, F., Bathke, A.C. & Pauly, M. (2020): Ranks and Pseudo-ranks - Surprising Results of Certain Rank Tests in Unbalanced Designs. *International Statistical Review*, doi:10.1111/insr.12418.

12. art1.anova: nonparametric analysis of variance using the ART (Aligned Rank Transform) for between subjects designs

factorial nonparametric analysis of variance.

Call: `art1.anova (model, dataframe, method=.., main=.., adjust=.., INT=..)`

Parameter:

model	anova model (as in function aov) example: x ~ A*B
dataframe	Data, object of type data.frame
method	0: Alignment using a regression (default) 1: Alignment by computing the deviations from the cell means
main	F: for tests of main effects use the simple RT-technique (default) T: for tests of main effects use also the ART-technique
INT	F: no transformation into normal scores after ranking (default) T: transformation into normal scores after ranking

Result:

anova table	object of type data.frame and anova

References:

Wobbrock, J. O., Findlater, L., Gergle, D. & Higgins, J. (2011): *The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only ANOVA Procedures*, Computer Human Interaction - CHI , pp. 143-146

Note:

Meanwhile there exists also the package ARTool on Cran for between subjects designs.

13. art2.anova: nonparametric analysis of variance using the ART (Aligned Rank Transform) for pure within subjects designs

1- and 2-factorial nonparametric analysis of variance for pure within subjects designs. For the tests of main effects the simple RT-technique is applied.

Call: `art2.anova (model, dataframe, main=..., INT=...)`

Parameter:

model	anova model (as in function aov) example: x ~ Medi*Aufgabe+Error(Vpn/(Medi*Aufgabe))
dataframe	Data, object of type data.frame (long format)
main	F: for tests of main effects use the simple RT-technique (default) T: for tests of main effects use also the ART-technique
INT	F: no transformation into normal scores after ranking (default) T: transformation into normal scores after ranking

Result:

anova table	object of type data.frame and anova

References:

Wobbrock, J. O., Findlater, L., Gergle, D. & Higgins, J. (2011): *The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only ANOVA Procedures*, Computer Human Interaction - CHI , pp. 143-146

14. art3.anova: nonparametric analysis of variance using the ART (Aligned Rank Transform) for mixed designs (split plot designs)

factorial nonparametric analysis of variance for mixed designs with at least one grouping factor and one or two repeated measurement factors. For the tests of main effects the simple RT-technique is applied.

Call: `art3.anova (`*`model, dataframe,`*` method=..., main=..., INT=...)`

Parameter:

model	anova model (as in function aov) example: score ~ gruppe*Zeit+Error(Vpn/Zeit)
dataframe	Data, object of type data.frame (long format)
method	0: Alignment using a regression (default) 1: Alignment by computing the deviations from the cell means
main	F: for tests of main effects use the simple RT-technique (default) T: for tests of main effects use also the ART-technique
INT	F: no transformation into normal scores after ranking (default) T: transformation into normal scores after ranking

Result:

anova table	object of type data.frame and anova

References:

Wobbrock, J. O., Findlater, L., Gergle, D. & Higgins, J. (2011): *The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only ANOVA Procedures*, Computer Human Interaction - CHI , pp. 143-146

15. koch.anova: nonparametric anova for split plot designs using the procedure by G. Koch

2-factorial nonparametric analysis of variance based on ranks for mixed designs (split plot designs) using the method of Gary Koch. This method does not assume spherecity of the covariance matrix of the dependent variables. But there are several alternatives for this method:

The grouping main effect can be tested with a multivariate Kruskal-Wallis test. It has to be pointed out to the fact that this test is not independent of the interaction effect, with the consequence that any interaction will lead to larger type I error rates for the test of the grouping effect. But in practice that is not severe because in the case of a significant interaction effect the tests of the main effects are no longer of interest. Alternatively the grouping main effect can be tested by means of a univariate Kruskal-Wallis test on the subject means (default).

For the test of the repeated measures main effect there exists a test $W_{ni}*$ assuming distributions which are equally shaped for all groups. Additionally, the sample sizes n_i should be larger than the number of repeated measurements. Here also the test is not independent of the interaction effect, with the consequence that any interaction will lead to larger type I error rates for the test of the repeated measurements effect. Alternatively there is a version of W_N* which is independent of the interaction at the cost of the power. On the other hand there exists a test W assuming arbitrary distribution shapes which is also unaffected by the interaction (default).

The dataframe must have the „wide“ format, i.e. all repeated measures variables must be in one row. The structure as it is requested by `aov` oder `ezANOVA` is not allowed. Missing values (NAs) have to be eliminated befor usage.

Call: `koch.anova (`*`dataframe, grouping factor,`*` A=.., B=..)`

Parameter:

dataframe	data, object of type data.frame
grouping factor	variable within the data.frame
A=	0: univariate Kruskal-Wallis test on the subject means 1: multivariate Kruskal-Wallis test
B=	0: W test for the entire sample, assuming arbitrary distribution shapes 1: W_N* test for the entire sample, assuming equally shaped distributions 2: $W_{ni}*$ test pooled over all groups, assuming equally shaped distribtns

Result:

anova table	object of type data.frame

Example:

```
koch.anova (mydata[ ,c("y1", "y2", "y3", "y4")], mydata$age)
```

References:

Gary Koch: *Some aspects of the statistical analysis of split plot experiments in completely randomized layouts*. Journal of the American Statistical Association, Vol. 64, No. 326 (Jun., 1969), pp. 485-505

16. iga and iga.anova: the general approximation test (GA) and the improved general approximation test (IGA) by H.Huynh

Adjustments for the parametric F test in 2-factorial split-plot designs as proposed by H. Huynh, to allow for nonspherical and heterogeneous covariance matrices. For the repeated measures main effect as well as for the interaction effect a correction factor for the F value and adjusted degrees of freedom are computed. The GA can be considered as an improvement of the well-known degrees of freedom adjustment by Huynh & Feldt, whereas the IGA as an improvement of the GA for the case of heterogeneous covariance matrices.

The function `iga` computes the correction factor and the adjusted degrees of freedom for the two effects, whereas the function `iga.anova` performs a complete split-plot anova combining the results with those from `iga` resulting in an adjusted anova table. `iga.anova` allows data either in long or in wide format.

Call: `iga (`*`dataframe, grouping factor`*`)`

Parameter:

dataframe	data of the dependent variables in wide format, object of class matrix or data.frame
grouping factor	vector of class factor

Result: list of 4 vectors, each having 3 elements

`GA.B`	GA-adjustments for the repeated measures main effect
`GA.AB`	GA-adjustments for the interaction effect
`IGA.B`	IGA-adjustments for the repeated measures main effect
`IGA.AB`	IGA-adjustments for the interaction effect

Each of the above vectors has 3 elements:

1. correction factor `c` for the F value

2. adjusted degrees of freedom for the numerator `df1`

3. adjusted degrees of freedom for the denominator `df2`

If F is the unadjusted F value from the parametric anova, then `cF` is to be tested with (`df1,df2`) degrees of freedom

Examples:

```
iga(winer518[,3:5],winer518,2])
```

Call (wide format): `iga.anova (`*`df, group, ga=F`*`)`

Call (long format): `with(`*`df`*`,iga.anova (`*`y, groups=g, trial=w, id=id, ga=F`*`)`

Parameters:

df	data frame either in wide format or in long format which contains in case of wide format only the dependent variables
y	dependent variable
g	grouping factor of class factor
w	repeated measures factor of class factor
id	case id variable of class factor
iga	F: IGA adjustment, T: GA adjustment

Examples:

```
iga.anova(winer518[,3:5],winer518,2])                # wide format
with(winer518t, iga.anova(score,Geschlecht,Zeit,Vpn))  # long format
```

References:

Huynh, H. (1978): *Some approximate tests for repeated measurement designs*, Psychometrika 43, pp 161-175.

17. ap.anova: nonparametric anova for repeated measures designs based on a multivariate test by Agresti & Pendergast

1- or 2-factorial nonparametric analysis of variance based on ranks for repeated measures designs (e.g. split plot designs) using a nonparametric multivariate test by A. Agresti & J. Pendergast. This method does not assume spherecity of the covariance matrix of the dependent variables. Only the tests of the repeated measures main effect and of the interaction are performed.

The dataframe must have the „long“ format, the standard for repeated measures designs. Missing values (NAs) have to be eliminated befor usage.

Call: `ap.anova (`*`dataframe, dependent var, case id, trial factor`*
`[,`*`grouping factor`*`])`

Parameter:

dataframe	data, object of class data.frame
dependent variable	dependent variable within the data.frame
case id	case identifier within the data.frame of class factor
trial factor	reapeated measures factor within the data.frame of class factor
grouping factor	grouping factor within the data.frame of class factor (optional)

Variable names must be stated in „...“.

Result:

anova table	object of class data.frame

Example:

```
ap.anova (winer518t,"score","Vpn","Geschlecht","Zeit")
```

References:

Beasley: Multivariate Aligned Rank Test for Interactions in Multiple Group Repeated Measures Designs, *Multivariate Behavioural Research*, 37, No. 2, pp 197-226.

Tian & Wilcox (2007): A Comparison of Two Rank Tests for Repeated Measures Designs, *Journal of Modern Applied Statistical Methods*, Vol. 6, No. 1, pp 331-335.

18. simple.effects: parametric analysis of simple effects for between subject and mixed designs

Parametric analysis of simple effects for designs with at least one grouping factor and at most one repeated measurement factors.

Call: `simple.effects (`*`anova, interaction, dataframe,`*` adjust=...)`

Parameter:

anova	anova result (as from function aov)
interaction	one or more interaction terms enclosed in “...“, e.g. `c("A*time","A*time")`
dataframe	dataframe which was used for the analysis by aov
adjust	optional: α-adjustment method (see R function `p.adjust`) default “none“, no adjustment

References:

B.J.Winer et al.: *Statistical Principles in Experimnetal Design*,
McGraw-Hill, New York, 1991, pp 422 and pp 526

19. gee.anova: Anova-like tests for GEE and GLMM models

There are 2 Anova-like Wald tests for 2-factorial designs: `gee.anova` for the classical Wald-test and `gee.robanova` for a robust Wald-test according to Fan & Zhang. The classical Wald test is rather liberal, especially in the case of GEE and GLMM models where the covariance matrix of the parameter estimates is generally underestimated, resulting in too large χ^2-values.

Call: `gee.anova` (*`coefficients, covariance matrix, degrees of freedom, n`*)

`gee.robanova` (*`coefficients, covariance matrix, degrees of freedom`*)

Parameter:

`oefficients`	regression coefficients (details see below)
`covariance matrix`	
`degrees of freedom`	Array with 3 df for 2 factors and the interaction
`n`	sample size (required for the F test)

Result:

The result is a dataframe with 3 rows, one for each of the 3 effects with columns:

`gee.anova:`
degrees of freedom
χ^2-value
p value

`gee.robanova`
degrees of freedom
χ^2-value
corresponding p value

F-value
corresponding p value

`nerror:` 0 for no errors
`err.invert:` 0 for no errors while computing the inverse

References:

Li, Peng & Redden, David T. (2015): Comparing denominator degrees of freedom approximations for the generalized linear mixed model in analyzing binary outcome in small sample cluster-randomized trials.
BMC Medical Research Methodology,
https://doi.org/10.1186/s12874-015-0026-x

Fan, C. & Zhang, D. (2014): Robust small sample inference for generalised estimating equations: An application of the Anova-type test.
Australian & New Zealand Journal of Statistics, 56(3), pp 237–255

Specification of the coefficients and covariance matrix:

In every model estimation the coefficients and covariance matrix are part of the resulting object. For the most popular of the functions to be used for the analyses the following table shows where these are to be found:

function	package	coefficients	covariance matrix
gls	nlme	...$coefficients	vcov(...) ...$varBeta
glm	stats	...$coefficients	vcov(...)
lmer	lme4	...@beta summary(...)$coefficients[,1]	vcov(...)
glmer	lme4	...@beta summary(...)$coefficients[,1]	vcov(...)
glmmML	glmmML	...$coefficients	...$variance
glmmPQL	MASS	,,,$coefficients$fixed	...$varFix
geeglm	geepack	...$coefficients	...$geese$vbeta
gee	drgee	...$coefficients	...$vcov
gee	gee	...$coefficients	...$"robust.variance"
wgee	wgeesel	...$beta	...$var
JGee1	JGEE	...$coefficients	...$"robust.variance"
MGEE	PGEE	...$coefficients	...$"robust.variance"