Package 'DFA.CANCOR'

January 3, 2022

Type Package

Title Linear Discriminant Function and Canonical Correlation Analysis

Version 0.2.3

Date 2022-01-03

Author Brian P. O'Connor

Maintainer Brian P. O'Connor <brian.oconnor@ubc.ca>

Description Produces SPSS- and SAS-like output for linear discriminant function analysis and canonical correlation analysis. The methods are described in Manly & Alberto (2017, ISBN:9781498728966), Rencher (2002, ISBN:0-471-41889-7), and Tabachnik & Fidell (2019, ISBN:9780134790541).

Imports graphics, stats, BayesFactor, MVN

LazyLoad yes

LazyData yes

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2022-01-03 13:40:12 UTC

R topics documented:

`DFA.CANCOR-package` *DFA.CANCOR*

Description

Provides SPSS- and SAS-like output for linear discriminant function analysis (via the DFA function) and for canonical correlation analysis (via the CANCOR function), and for providing effect sizes and significance tests for pairwise group comparisons (via the GROUP.DIFFS function). There are also functions for assessing the assumptions of normality, linearity, and homogeneity of variances and covariances.

`CANCOR` *Canonical correlation analysis*

Description

Produces SPSS- and SAS-like output for canonical correlation analysis. Portions of the code were adapted from James Steiger (www.statpower.net).

Usage

```
CANCOR(data, set1, set2, plot, plotCV, plotcoefs, verbose)
```

Arguments

`data`	A dataframe where the rows are cases & the columns are the variables.
`set1`	The names of the continuous variables for the first set, e.g., set1 = c('varA', 'varB', 'varC').
`set2`	The names of the continuous variables for the second set, e.g., set2 = c('varD', 'varE', 'varF').
`plot`	Should a plot of the coefficients be produced? The options are: TRUE (default) or FALSE.
`plotCV`	The canonical variate number for the plot, e.g., plotCV = 1.
`plotcoefs`	The coefficient for the plots. The options are 'structure' (default) or 'standardized'.
`verbose`	Should detailed results be displayed in the console? The options are: TRUE (default) or FALSE.

Value

If verbose = TRUE, the displayed output includes Pearson correlations, multivariate significance tests, canonical function correlations and bivariate significance tests, raw canonical coefficients, structure coefficients, standardized coefficients, and a bar plot of the structure or standardized coefficients.

The returned output is a list with elements

`cancorrels`	canonical correlations and their significance tests
`CoefRawSet1`	raw canonical coefficients for Set 1
`CoefRawSet2`	raw canonical coefficients for Set 2
`CoefStruct11`	structure coefficients for Set 1 variables with the Set 1 variates
`CoefStruct21`	structure coefficients for Set 2 variables with the Set 1 variates
`CoefStruct12`	structure coefficients for Set 1 variables with the Set 2 variates
`CoefStruct22`	structure coefficients for Set 2 variables with the Set 2 variates
`CoefStandSet1`	standardized coefficients for Set 1 variables
`CoefStandSet2`	standardized coefficients for Set 2 variables
`mv_Wilk`	Wilk's multivariate significance test
`mv_Pillai`	Pillai-Bartlett multivariate significance test
`mv_Hotelling`	Hotelling-Lawley multivariate significance test
`mv_Roy`	Roy's Largest Root multivariate significance test
`mv_BartlettV`	Bartlett's V multivariate significance test
`mv_Rao`	Rao's' multivariate significance test
`CorrelSet1`	Pearson correlations for Set 1
`CorrelSet2`	Pearson correlations for Set 2
`CorrelSet1n2`	Pearson correlations between Set 1 & Set 2

Author(s)

Brian P. O'Connor

References

Manly, B. F. J., & Alberto, J. A. (2017). *Multivariate statistical methods: A primer (4th Edition).* Chapman & Hall/CRC, Boca Raton, FL.

Rencher, A. C. (2002). *Methods of Multivariate Analysis* (2nd ed.). New York, NY: John Wiley & Sons.

Sherry, A., & Henson, R. K. (2005). Conducting and interpreting canonical correlation analysis in personality research: A user-friendly primer. *Journal of Personality Assessment, 84,* 37-48.

Steiger, J. (2019). *Canonical correlation analysis.*
www.statpower.net/Content/312/Lecture%20Slides/CanonicalCorrelation.pdf

Tabachnik, B. G., & Fidell, L. S. (2019). *Using multivariate statistics (7th ed.).* New York, NY: Pearson.

Examples

```
# data that simulate those from De Leo & Wulfert (2013)
CANCOR(data = na.omit(data_CCA_De_Leo),
       set1 = c('Tobacco_Use','Alcohol_Use','Illicit_Drug_Use','Gambling_Behavior',
                'Unprotected_Sex','CIAS_Total'),
       set2 = c('Impulsivity','Social_Interaction_Anxiety','Depression',
                'Social_Support','Intolerance_of_Deviance','Family_Morals',
                'Family_Conflict','Grade_Point_Average'),
       plot = TRUE, plotCV = 1, plotcoefs='structure',
       verbose = TRUE)


# data from Tabachnik & Fidell (2013, p. 589)
CANCOR(data = data_CCA_Tabachnik,
       set1 = c('TS','TC'),
       set2 = c('BS','BC'),
       plot = TRUE, plotCV = 1, plotcoefs='structure',
       verbose = TRUE)

# UCLA dataset
UCLA_CCA_data <- read.csv("https://stats.idre.ucla.edu/stat/data/mmreg.csv")
colnames(UCLA_CCA_data) <- c("LocusControl", "SelfConcept", "Motivation",
                             "read", "write", "math", "science", "female")
summary(UCLA_CCA_data)
CANCOR(data = UCLA_CCA_data,
       set1 = c("LocusControl","SelfConcept","Motivation"),
       set2 = c("read","write","math","science","female"),
       plot = TRUE, plotCV = 1, plotcoefs='standardized',
       verbose = TRUE)
```

---

data_CCA_De_Leo *data_CCA_De_Leo*

Description

A data frame with scores on 14 variables that have the same correlational structure, and which produce the same canonical correlation analysis results, as those reported in De Leo and Wulfert (2013).

Usage

```
data(data_CCA_De_Leo)
```

Source

De Leo, J. A., & Wulfert, E. (2013). Problematic internet use and other risky behaviors in college students: An application of problem-behavior theory. *Psychology of Addictive Behaviors, 27(1),* 133-141.

Examples

```
head(data_CCA_De_Leo)

CANCOR(data = na.omit(data_CCA_De_Leo),
       set1 = c('Tobacco_Use','Alcohol_Use','Illicit_Drug_Use','Gambling_Behavior',
                'Unprotected_Sex','CIAS_Total'),
       set2 = c('Impulsivity','Social_Interaction_Anxiety','Depression',
                'Social_Support','Intolerance_of_Deviance','Family_Morals',
                'Family_Conflict','Grade_Point_Average'),
       plot = 'yes', plotCV = 1,
       verbose=TRUE)
```

`data_CCA_Tabachnik` *data_CCA_Tabachnik*

Description

A data frame with scores on 4 variables for 8 cases. Used by Tabachnik & Fidell (2019, p. 451) in their chapter on canonical correlation.

Usage

```
data(data_CCA_Tabachnik)
```

Source

Tabachnik, B. G., & Fidell, L. S. (2019). *Using multivariate statistics (7th ed.).* New York, NY: Pearson.

Examples

```
head(data_CCA_Tabachnik)

CANCOR(data = data_CCA_Tabachnik,
       set1 = c('TS','TC'),
       set2 = c('BS','BC'),
       plot = 'yes', plotCV = 1,
       verbose=TRUE)
```

data_DFA_Field *data_DFA_Field*

Description

A data frame with scores on 2 variables for 10 cases in each of 3 groups. Used by Field et al. (2012) in their chapter on MANOVA and discriminant function analysis.

Usage

```
data(data_DFA_Field)
```

Source

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R.* Los Angeles, CA: Sage.

Examples

```
head(data_DFA_Field)

DFA(data = data_DFA_Field,
    groups = 'Group',
    variables = c('Actions','Thoughts'),
    predictive = TRUE, priorprob = 'SIZES',
    verbose = TRUE)
```

data_DFA_Sherry *data_DFA_Sherry*

Description

A data frame with scores on 5 variables for 10 cases in each of 3 groups. Used by Sherry (2006) in her article on discriminant function analysis.

Usage

```
data(data_DFA_Sherry)
```

Source

Sherry, A. (2006). Discriminant analysis in counseling research. *Counseling Psychologist, 34,* 661-683.

Examples

```
head(data_DFA_Sherry)

DFA(data = data_DFA_Sherry,
    groups = 'Group',
    variables = c('Neuroticism','Extroversion','Openness',
                  'Agreeableness','Conscientiousness'),
    predictive = TRUE, priorprob = 'SIZES',
    verbose=TRUE)
```

DFA *Discriminant function analysis*

Description

Produces SPSS- and SAS-like output for linear discriminant function analysis.

Usage

```
DFA(data, groups, variables, plot, predictive, priorprob, covmat_type, CV, verbose)
```

Arguments

`data`	A dataframe where the rows are cases & the columns are the variables.
`groups`	The name of the groups variable in the dataframe, e.g., groups = 'Group'.
`variables`	The names of the continuous variables in the dataframe that will be used in the DFA, e.g., variables = c('varA', 'varB', 'varC').
`plot`	Should a plot of the mean standardized discriminant function scores for the groups be produced? The options are: TRUE (default) or FALSE.
`predictive`	Should a predictive DFA be conducted? The options are: TRUE (default) or FALSE.
`priorprob`	If predictive = TRUE, how should the prior probabilities of the group sizes be computed? The options are: 'EQUAL' for equal group sizes; or 'SIZES' (default) for the group sizes to be based on the sizes of the groups in the dataframe.
`covmat_type`	The kind of covariance to be used for a predictive DFA. The options are: 'within' (for the pooled within-groups covariance matrix, which is the default) or 'separate' (for separate-groups covariance matrices).
`CV`	If predictive = TRUE, should cross-validation (leave-one-out cross-validation) analyses also be conducted? The options are: TRUE (default) or FALSE.
`verbose`	Should detailed results be displayed in console? The options are: TRUE (default) or FALSE.

Details

The predictive DFA option using separate-groups covariance matrices (which is often called 'quadratic DFA') is conducted following the procedures described by Rencher (2002). The covariance matrices in this case are based on the scores on the continuous variables. In contrast, the 'separate-groups' option in SPSS involves use of the group scores on the discriminant functions (not the original continuous variables), which can produce different classifications.

See the documentation below for the GROUP.DIFFS function for information on the interpretation of the Bayes factors and effect sizes that are produced for the group comparisons.

Value

If verbose = TRUE, the displayed output includes descriptive statistics for the groups, tests of univariate and multivariate normality, the results of tests of the homogeneity of the group variance-covariance matrices, eigenvalues & canonical correlations, Wilks lambda & peel-down statistics, raw and standardized discriminant function coefficients, structure coefficients, functions at group centroids, one-way ANOVA tests of group differences in scores on each discriminant function, one-way ANOVA tests of group differences in scores on each original DV, significance tests for group differences on the original DVs according to Bird et al. (2014), a plot of the group means on the standardized discriminant functions, and extensive output from predictive discriminant function analyses (if requested).

The returned output is a list with elements

`rawCoef` canonical discriminant function coefficients

`structCoef` structure coefficients

`standCoef` standardized coefficients

`standCoefSPSS` standardized coefficients from SPSS

`centroids` unstandardized canonical discriminant functions evaluated at the group means

`centroidSDs` group standard deviations on the unstandardized functions

`centroidsZ` standardized canonical discriminant functions evaluated at the group means

`centroidSDsZ` group standard deviations on the standardized functions

`DFAscores` scores on the discriminant functions

`anovaDFoutput` One-way ANOVAs using the scores on a discriminant function as the DV

`anovaDVoutput` One-way ANOVAs on the original DVs

`MFWER1.sigtest` Significance tests when controlling the MFWER by (only) carrying out multiple t tests

`MFWER2.sigtest` Significance tests for the two-stage approach to controling the MFWER

`dfa_class` The predicted group classifications

`posteriors` The posterior probabilities for the predicted group classifications

`freqs_OR` Cross-tabulation of the original and predicted group memberships

`PropOrigCorrect` Proportion of original grouped cases correctly classified

`chi_square_OR` Chi-square test of independence

PressQ_OR	Press's Q significance test of classifiation accuracy for original vs. predicted group memberships
rowfreqs_OR	Row Frequencies
colfreqs_OR	Column Frequencies
cellprops_OR	Cell Proportions
rowprops_OR	Row-Based Proportions
colprops_OR	Column-Based Proportions
kappas_cvo_OR	Agreement (kappas) between the predicted and original group memberships
dfa_class_CV	Classifications from leave-one-out cross-validations
freqs_CV	Cross-Tabulation of the cross-validated and predicted group memberships
PropCrossValCorrect	Proportion of cross-validated grouped cases correctly classified
chi_square_CV	Chi-square test of indepedence
PressQ_CV	Press's Q significance test of classifiation accuracy for cross-validated vs. predicted group memberships
rowfreqs_CV	Row frequencies
colfreqs_CV	Column frequencies
cellprops_CV	Cell proportions
rowprops_CV	Row-based proportions
colprops_CV	Column-based proportions
kappas_cvoCV	Agreement (kappas) between the cross-validated and original group memberships
kappas_CVP	Agreement (kappas) between the cross-validated and predicted group memberships

Author(s)

Brian P. O'Connor

References

Bird, K. D., & Hadzi-Pavlovic, D. (2013). Controlling the maximum familywise Type I error rate in analyses of multivariate experiments. *Psychological Methods, 19(2),* p. 265-280.

Manly, B. F. J., & Alberto, J. A. (2017). *Multivariate statistical methods: A primer (4th Edition).* Chapman & Hall/CRC, Boca Raton, FL.

Rencher, A. C. (2002). *Methods of Multivariate Analysis* (2nd ed.). New York, NY: John Wiley & Sons.

Sherry, A. (2006). Discriminant analysis in counseling research. *Counseling Psychologist, 34,* 661-683.

Tabachnik, B. G., & Fidell, L. S. (2019). *Using multivariate statistics (7th ed.).* New York, NY: Pearson.

Examples

```
DFA_Field=DFA(data = data_DFA_Field,
    groups = 'Group',
    variables = c('Actions','Thoughts'),
    predictive = TRUE, priorprob = 'SIZES',
    covmat_type='separate', # altho better to used 'separate' for these data
    verbose = TRUE)



# plots of posterior probabilities by group
# hoping to see correct separations between cases from different groups

# first, display the posterior probabilities
print(cbind(round(DFA_Field$posteriors[1:3],3), DFA_Field$posteriors[4]))

# group NT vs CBT
plot(DFA_Field$posteriors$posterior_NT, DFA_Field$posteriors$posterior_CBT,
     pch = 16, col = c('red', 'blue', 'green')[DFA_Field$posteriors$Group],
     xlim=c(0,1), ylim=c(0,1),
     main = 'DFA Posterior Probabilities by Original Group Memberships',
     xlab='Posterior Probability of Being in Group NT',
     ylab='Posterior Probability of Being in Group CBT' )
legend(x=.8, y=.99, c('CBT','BT','NT'), cex=1.2, col=c('red', 'blue', 'green'), pch=16, bty='n')

# group NT vs BT
plot(DFA_Field$posteriors$posterior_NT, DFA_Field$posteriors$posterior_BT,
     pch = 16, col = c('red', 'blue', 'green')[DFA_Field$posteriors$Group],
     xlim=c(0,1), ylim=c(0,1),
     main = 'DFA Posterior Probabilities by Group Membership',
     xlab='Posterior Probability of Being in Group NT',
     ylab='Posterior Probability of Being in Group BT' )
legend(x=.8, y=.99, c('CBT','BT','NT'), cex=1.2,col=c('red', 'blue', 'green'), pch=16, bty='n')

# group CBT vs BT
plot(DFA_Field$posteriors$posterior_CBT, DFA_Field$posteriors$posterior_BT,
     pch = 16, col = c('red', 'blue', 'green')[DFA_Field$posteriors$Group],
     xlim=c(0,1), ylim=c(0,1),
     main = 'DFA Posterior Probabilities by Group Membership',
     xlab='Posterior Probability of Being in Group CBT',
     ylab='Posterior Probability of Being in Group BT' )
legend(x=.8, y=.99, c('CBT','BT','NT'), cex=1.2, col=c('red', 'blue', 'green'), pch=16, bty='n')


DFA_Sherry <- DFA(data = data_DFA_Sherry,
                  groups = 'Group',
                  variables = c('Neuroticism','Extroversion','Openness',
                                'Agreeableness','Conscientiousness'),
                  predictive = TRUE, priorprob = 'SIZES',
                  covmat_type='separate',
                  verbose = TRUE)
```

```
# plots of posterior probabilities by group
# hoping to see correct separations between cases from different groups

# first, display the posterior probabilities
print(cbind(round(DFA_Sherry$posteriors[1:3],3), DFA_Sherry$posteriors[4]))

# group 1 vs 2
plot(DFA_Sherry$posteriors$posterior_1, DFA_Sherry$posteriors$posterior_2,
     pch = 16, cex = 1, col = c('red', 'blue', 'green')[DFA_Sherry$posteriors$Group],
     xlim=c(0,1), ylim=c(0,1),
     main = 'DFA Posterior Probabilities by Original Group Memberships',
     xlab='Posterior Probability of Being in Group 1',
     ylab='Posterior Probability of Being in Group 2' )
legend(x=.8, y=.99, c('1','2','3'), cex=1.2, col=c('red', 'blue', 'green'), pch=16, bty='n')

# group 1 vs 3
plot(DFA_Sherry$posteriors$posterior_1, DFA_Sherry$posteriors$posterior_3,
     pch = 16, col = c('red', 'blue', 'green')[DFA_Sherry$posteriors$Group],
     xlim=c(0,1), ylim=c(0,1),
     main = 'DFA Posterior Probabilities by Group Membership',
     xlab='Posterior Probability of Being in Group 1',
     ylab='Posterior Probability of Being in Group 3' )
legend(x=.8, y=.99, c('1','2','3'), cex=1.2,col=c('red', 'blue', 'green'), pch=16, bty='n')

# group 2 vs 3
plot(DFA_Sherry$posteriors$posterior_2, DFA_Sherry$posteriors$posterior_3,
     pch = 16, col = c('red', 'blue', 'green')[DFA_Sherry$posteriors$Group],
     xlim=c(0,1), ylim=c(0,1),
     main = 'DFA Posterior Probabilities by Group Membership',
     xlab='Posterior Probability of Being in Group 2',
     ylab='Posterior Probability of Being in Group 3' )
legend(x=.8, y=.99, c('1','2','3'), cex=1.2, col=c('red', 'blue', 'green'), pch=16, bty='n')



# data from Tabachnik & Fiddel (2019, p 307)
table9.1 <- '
1  87  5 31 6.4
1  97  7 36 8.3
1 112  9 42 7.2
2 102 16 45 7.0
2  85 10 38 7.6
2  76  9 32 6.2
3 120 12 30 8.4
3  85  8 28 6.3
3  99  9 27 8.2'
table9.1 <- data.frame(read.table(text=table9.1,
                       col.names=c('group','perf','info','verbexp','age')))

DFA(data = table9.1,
    groups = 'group',
    variables = c('perf','info','verbexp','age'),
    predictive = TRUE, priorprob = 'SIZES', covmat_type='within',
```

```
verbose = TRUE)
```

GROUP.DIFFS *Group Mean Differences on a Continuous Outcome Variable*

Description

Produces a variety of statistics for all possible pairwise independent groups comparisons of means on a continuous outcome variable.

Usage

```
GROUP.DIFFS(data, GROUPS=NULL, DV=NULL, var.equal=FALSE, p.adjust.method="holm",
            Ncomps=NULL, verbose=TRUE)
```

Arguments

data	A dataframe where the rows are cases & the columns are the variables. If GROUPS and DV are not specified, then the GROUPS variable should be in the first column and the DV should be in the second column of data.
GROUPS	The name of the groups variable in the dataframe, e.g., groups = 'Group'.
DV	The name of the dependent (outcome) variable in the dataframe, e.g., DV = 'esteem'.
var.equal	(from stats::t.test) A logical variable indicating whether to treat the two variances as being equal. If TRUE then the pooled variance is used to estimate the variance otherwise the Welch (or Satterthwaite) approximation to the degrees of freedom is used.
p.adjust.method	The method to be used to adjust the p values for the number of comparisons. The options are "holm" (the default), "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none".
Ncomps	The number of pairwise comparisons for the adjusted p values. If unspecified, it will be the number of all possible comparisons (i.e., the family-wise number of number of comparisons). Ncomps could alternatively be set to, e.g., the experiment-wise number of number of comparisons.
verbose	Should detailed results be displayed in console? The options are: TRUE (default) or FALSE.

Details

The function conducts all possible pairwise comparisons of the levels of the GROUPS variable on the continuous outcome variable. It supplements independent groups t-test results with effect size statistics and with the Bayes factor for each pairwise comparison.

The d values are the Cohen d effect sizes, i.e., the mean difference expressed in standard deviation units.

The g values are the Hedges g value corrections to the Cohen d effect sizes.

The r values are the effect sizes for the group mean difference expressed in the metric of Pearson's r.

The BESD values are the binomial effect size values for the group mean differences. The BESD casts the effect size in terms of the success rate for the implementation of a hypothetical procedure (e.g., the percentage of cases that were cured, or who died.) For example, an r = .32 is equivalent to increasing the success rate from 34% to 66% (or, possibly, reducing an illness or death rate from 66% to 34%).

The Bayes factor values are obtained from the ttest.tstat function in the BayesFactor package.

For example, a Bayes_Factor_alt_vs_null = 3 indicates that the data are 3 times *more* likely under the alternative hypothesis than under the null hypothesis. A Bayes_Factor_alt_vs_null = .2 indicates that the data are five times *less* likely under the alternative hypothesis than under the null hypothesis (1 / .2).

Conversely, a Bayes_Factor_null_vs_alt = 3 indicates that the data are 3 times *more* likely under the null hypothesis than under the alternative hypothesis. A Bayes_Factor_null_vs_alt = .2 indicates that the data are five times *less* likely under the null hypothesis than under the alternative hypothesis (1 / .2).

Value

If verbose = TRUE, the displayed output includes the means, standard deviations, and Ns for the groups, the t-test results for each pairwise comparison, the mean difference and its 95% confidence interval, four indices of effect size for each pairwise comparison (r, d, g, and BESD), and the Bayes factor. The returned output is a matrix with these values.

Author(s)

Brian P. O'Connor

References

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science, 2(2),* 156168.

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *Journal of Problem Solving, 7,* 29.

Randolph, J. & Edmondson, R.S. (2005). Using the binomial effect size display (BESD) to present the magnitude of effect sizes to the evaluation audience. *Practical Assessment Research & Evaluation, 10,* 14.

Rosenthal, R., Rosnow, R.L., & Rubin, D.R. (2000). *Contrasts and effect sizes in behavioral research: A correlational approach.* Cambridge UK: Cambridge University Press.

Rosenthal, R., & Rubin, D. B. (1982). A simple general purpose display of magnitude and experimental effect. *Journal of Educational Psychology, 74,* 166-169.

Rouder, J. N., Haaf, J. M., & Vandekerckhove, J. (2018). Bayesian inference for psychology, part IV: parameter estimation and Bayes factors. *Psychonomic Bulletin & Review, 25(1),* 102113.

Examples

```
GROUP.DIFFS(data_DFA_Field, var.equal=FALSE, p.adjust.method="fdr")


GROUP.DIFFS(data = data_DFA_Sherry, var.equal=FALSE, p.adjust.method="bonferroni")
```

`HOMOGENEITY` *Homogeneity of variances and covariances*

Description

Produces tests of the homogeneity of variances and covariances.

Usage

```
HOMOGENEITY(data, groups, variables, verbose)
```

Arguments

`data`	A dataframe where the rows are cases & the columns are the variables.
`groups`	(optional) The name of the groups variable in the dataframe (if there is one) e.g., groups = 'Group'.
`variables`	(optional) The names of the continuous variables in the dataframe for the analyses, e.g., variables = c('varA', 'varB', 'varC').
`verbose`	Should detailed results be displayed in the console? The options are: TRUE (default) or FALSE.

Value

If "variables" is specified, the analyses will be run on the "variables" in "data". If verbose = TRUE, the displayed output includes descriptive statistics and tests of univariate and multivariate homogeneity.

Bartlett's test compares the variances of k samples. The data must be normally distributed.

The non-parametric Fligner-Killeen test also compares the variances of k samples and it is robust when there are departures from normality.

Box's M test is a multivariate statistical test of the equality of multiple variance-covariance matrices. The test is prone to errors when the sample sizes are small or when the data do not meet model assumptions, especially the assumption of multivariate normality. For large samples, Box's M test may be too strict, indicating heterogeneity when the covariance matrices are not very different.

The returned output is a list with elements

covmatrix	The variance-covariance matrix for each group
Bartlett	Bartlett test of homogeneity of variances (parametric)
Figner_Killeen	Figner-Killeen test of homogeneity of variances (non parametric)
PooledWithinCovarSPSS	the pooled within groups covariance matrix from SPSS
PooledWithinCorrelSPSS	the pooled within groups correlation matrix from SPSS
sscpWithin	the within sums of squares and cross-products matrix
sscpBetween	the between sums of squares and cross-products matrix
BoxLogdets	the log determinants for Box's test
BoxMtest	Box's' test of the equality of covariance matrices

Author(s)

Brian P. O'Connor

References

Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika, 36 (3-4),* 317-346.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London Series A 160,* 268-282.

Conover, W. J., Johnson, M. E., & Johnson, M. M. (1981). A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics, 23,* 351-361.

Warner, R. M. (2013). *Applied statistics: From bivariate through multivariate techniques.* Thousand Oaks, CA: SAGE.

Examples

```
# data from Field et al. (2012)
HOMOGENEITY(data = data_DFA_Field,
            groups = 'Group', variables = c('Actions','Thoughts'))


# data from Sherry (2006)
HOMOGENEITY(data = data_DFA_Sherry,
            groups = 'Group',
            variables = c('Neuroticism','Extroversion','Openness',
                          'Agreeableness','Conscientiousness'))
```

LINEARITY *Linearity*

Description

Provides tests of the possible linear and quadratic associations between two continuous variables.

Usage

```
LINEARITY(data, variables, groups, idvs, dv, verbose)
```

Arguments

`data`	A dataframe where the rows are cases & the columns are the variables.
`variables`	(optional) The names of the continuous variables in the dataframe for the analyses, e.g., variables = c('varA', 'varB', 'varC').
`groups`	(optional) The name of the groups variable in the dataframe (if there is one), e.g., groups = 'Group'.
`idvs`	(optional) The names of the predictor variables, e.g., variables = c('varA', 'varB', 'varC').
`dv`	(optional) The name of the dependent variable, if output for just one dependent variable is desired.
`verbose`	(optional) Should detailed results be displayed in the console? The options are: TRUE (default) or FALSE.

Value

If "variables" is specified, the analyses will be run on the "variables" in "data". If "groups" is specified, the analyses will be run for every value of "groups". If verbose = TRUE, the linear and quadratic regression coefficients and their statistical tests are displayed.

The returned output is a list with the regression coefficients and their statistical tests.

Author(s)

Brian P. O'Connor

References

Tabachnik, B. G., & Fidell, L. S. (2019). *Using multivariate statistics (7th ed.).* New York, NY: Pearson.

Examples

```
# data from Sherry (2006), using all variables
LINEARITY(data=data_DFA_Sherry, groups='Group',
          variables=c('Neuroticism','Extroversion','Openness',
                      'Agreeableness','Conscientiousness') )


# data from Sherry (2006), specifying independent variables and a dependent variable
LINEARITY(data=data_DFA_Sherry, groups='Group',
          idvs=c('Neuroticism','Extroversion','Openness','Agreeableness'),
          dv=c('Conscientiousness'),
          verbose=TRUE )

# data that simulate those from De Leo & Wulfert (2013)
LINEARITY(data=data_CCA_De_Leo,
          variables=c('Tobacco_Use','Alcohol_Use','Illicit_Drug_Use',
                      'Gambling_Behavior', 'Unprotected_Sex','CIAS_Total',
                      'Impulsivity','Social_Interaction_Anxiety','Depression',
                      'Social_Support','Intolerance_of_Deviance','Family_Morals',
                      'Family_Conflict','Grade_Point_Average'),
          verbose=TRUE )
```

`NORMALITY` *Univariate and multivariate normality*

Description

Produces tests of univariate and multivariate normality using the MVN package.

Usage

```
NORMALITY(data, groups, variables, verbose)
```

Arguments

`data`	A dataframe where the rows are cases & the columns are the variables.
`groups`	(optional) The name of the groups variable in the dataframe, e.g., groups = 'Group'.
`variables`	(optional) The names of the continuous variables in the dataframe for the analyses, e.g., variables = c('varA', 'varB', 'varC').
`verbose`	Should detailed results be displayed in the console? The options are: TRUE (default) or FALSE.

Value

If "groups" is not specified, the analyses will be run on all of the variables in "data". If "groups" is specified, the analyses will be run for every value of "groups". If "variables" is specified, the analyses will be run on the "variables" in "data". If verbose = TRUE, the displayed output includes descriptive statistics and tests of univariate and multivariate normality.

The returned output is a list with elements

`descriptives`	descriptive statistics, including skewness and kurtosis
`Shapiro_Wilk`	the Shapiro_Wilk test of univariate normality
`Mardia`	the Mardia test of multivariate normality
`Henze_Zirkler`	the Henze-Zirkler test of multivariate normality
`Royston`	the Royston test of multivariate normality
`Doornik_Hansen`	the Doornik_Hansen test of multivariate normality

Author(s)

Brian P. O'Connor

References

Korkmaz, S., Goksuluk, D., Zararsiz, G. (2014). MVN: An R package for assessing multivariate normality. *The R Journal, 6(2),* 151-162.

Szekely,G. J., & Rizzo, M. L. (2017). The energy of data. *Annual Review of Statistics and Its Application 4,* 447-79.

Tabachnik, B. G., & Fidell, L. S. (2019). *Using multivariate statistics (7th ed.).* New York, NY: Pearson.

Examples

```
# data that simulate those from De Leo & Wulfert (2013)
NORMALITY(data = na.omit(data_CCA_De_Leo[c(
          'Unprotected_Sex','Tobacco_Use','Alcohol_Use','Illicit_Drug_Use',
          'Gambling_Behavior','CIAS_Total','Impulsivity','Social_Interaction_Anxiety',
          'Depression','Social_Support','Intolerance_of_Deviance','Family_Morals',
          'Family_Conflict','Grade_Point_Average')]))


# data from Field et al. (2012)
NORMALITY(data = data_DFA_Field,
          groups = 'Group',
          variables = c('Actions','Thoughts'))

# data from Tabachnik & Fidell (2013, p. 589)
NORMALITY(data = na.omit(data_CCA_Tabachnik[c('TS','TC','BS','BC')]))

# UCLA dataset
UCLA_CCA_data <- read.csv("https://stats.idre.ucla.edu/stat/data/mmreg.csv")
```

```
colnames(UCLA_CCA_data) <- c("LocusControl", "SelfConcept", "Motivation",
                             "read", "write", "math", "science", "female")
summary(UCLA_CCA_data)
NORMALITY(data = na.omit(UCLA_CCA_data[c("LocusControl","SelfConcept","Motivation",
                                          "read","write","math","science","female")]))
```

PLOT_LINEARITY *Plot for linearity*

Description

Plots the linear, quadratic, and loess regression lines for the association between two continuous variables.

Usage

```
PLOT_LINEARITY(data, idv, dv, groups=NULL, groupNAME=NULL, legposition=NULL,
               leginset=NULL, verbose=TRUE)
```

Arguments

`data`	A dataframe where the rows are cases & the columns are the variables.
`idv`	The name of the predictor variable.
`dv`	The name of the dependent variable.
`groups`	(optional) The name of the groups variable in the dataframe, e.g., groups = 'Group'.
`groupNAME`	(optional) The value (level, name, or number) from the groups variable that identifies the subset group whose data will be used for the analyses, e.g., groupNAME = 1.
`legposition`	(optional) The position of the legend, as specified by one of the following possible keywords: "bottomright", "bottom", "bottomleft", "left", "topleft", "top", "topright", "right" or "center".
`leginset`	(optional) The inset distance(s) of the legend from the margins as a fraction of the plot region when legend is placed by keyword.
`verbose`	Should detailed results be displayed in the console? The options are: TRUE (default) or FALSE.

Value

If verbose = TRUE, the linear and quadratic regression coefficients and their statistical tests are displayed.

The returned output is a list with the regression coefficients and the plot data.

Author(s)

Brian P. O'Connor

References

Tabachnik, B. G., & Fidell, L. S. (2019). *Using multivariate statistics (7th ed.).* New York, NY: Pearson.

Examples

```
# data that simulate those from De Leo & Wulfert (2013)
PLOT_LINEARITY(data=data_CCA_De_Leo, groups=NULL,
               idv='Family_Conflict', dv='Grade_Point_Average', verbose=TRUE)


# data from Sherry (2006), ignoring the groups
PLOT_LINEARITY(data=data_DFA_Sherry, groups=NULL, groupNAME=NULL,
               idv='Neuroticism', dv='Conscientiousness', verbose=TRUE)

# data from Sherry (2006), group 2 only
PLOT_LINEARITY(data=data_DFA_Sherry, groups ='Group', groupNAME=2,
               idv='Neuroticism', dv='Conscientiousness', verbose=TRUE)
```

Index