Package ‘DFA.CANCOR’

May 6, 2020

Type Package
Title Linear Discriminant Function and Canonical Correlation Analysis
Version 0.1.8
Date 2020-05-3
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
Tabachnik & Fidell (2013, ISBN-10:0-205-89081-4), and
Imports MASS, MVN, graphics, stats
LazyLoad yes
LazyData yes
License GPL (>= 2)
NeedsCompilation no
Repository CRAN
Date/Publication 2020-05-06 08:40:02 UTC

R topics documented:

DFA.CANCOR-package .................................................. 2
CANCOR ................................................................. 2
data_CCA_De_Leo ...................................................... 4
data_CCA_Tabachnik .................................................... 5
data_DFA_Field .......................................................... 6
data_DFA_Sherry ......................................................... 6
DFA ........................................................................... 7
homogeneity ............................................................... 10
linearity ................................................................. 12
normality ............................................................... 13
plot_linearity .......................................................... 15
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). There are also functions for assessing the assumptions of normality, linearity, and homogeneity of variances and covariances.

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


www.statpower.net/Content/312/Lecture%20Slides/CanonicalCorrelation.pdf

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",
               "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

**Examples**

```r
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)
```

---

**Description**

A data frame with scores on 4 variables for 8 cases. Used by Tabachnik & Fidell (2013, p. 589) in their chapter on canonical correlation.

**Usage**

```r
data(data_CCA_Tabachnik)
```

**Source**


**Examples**

```r
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**

**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**

```r
data(data_DFA_Field)
```

**Source**


**Examples**

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

---

**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**

```r
data(data_DFA_Sherry)
```

**Source**

### DFA

**Examples**

```r
head(data_DFA_Sherri)
DFA(data = data_DFA_Sherri,
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. It uses functions from the MASS package.

**Usage**

`DFA(data, groups, variables, plot, predictive, priorprob, 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.

- **verbose**  
  Should detailed results be displayed in console? The options are: TRUE (default) or FALSE.
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 tests
- `MFWER2.sigtest`: Significance tests for the two-stage approach to controlling the MFWER
- `ldaoutputCV`: Classifications from leave-one-out cross-validations
- `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 classification 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
- `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 independence
PressQ_CV           Press’s Q significance test of classification 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 Membership
kappas_CVP          Agreement (kappas) between the Cross-Validated and Predicted Group Memberships

Author(s)
Brian P. O’Connor

References


Examples
```
DFA(data = data_DFA_Field,
    groups = 'Group',
    variables = c('Actions','Thoughts'),
    predictive = TRUE, priorprob = 'SIZES',
    verbose = TRUE)
```
```
DFA(data = data_DFA_Sherry,
    groups = 'Group',
```
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 analy-
            ses, 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 homo-
ogeneity.

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)
homogeneity

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


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'))
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

normality

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


Examples

```r
# 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", ...)"
plot_linearity

```
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, 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.

- **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**

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

CANCOR, 2

data_CCA_De_Leo, 4
data_CCA_Tabachnik, 5
data_DFA_Field, 6
data_DFA_Sherry, 6
DFA, 7
DFA.CANCOR-package, 2

homogeneity, 10
linearity, 12
normality, 13
plot_linearity, 15