Package 'caresid'

June 15, 2023

Title Correspondence Analysis Plot and Associations Visualisation

Version 0.1

Description Performs a Correspondence Analysis (CA) on a contingency table and creates a scatter-plot of the row and column points on the selected dimensions. Optionally, the function can add segments to the plot to visualize significant associations between row and column categories on the basis of positive (unadjusted) standardized residuals larger than a given threshold.

Depends R (>= 4.0.0)

Imports ca (>= 0.71), ggplot2 (>= 3.4.0), ggrepel (>= 0.9.0), graphics (>= 4.2.0), stats (>= 4.2.0)

License GPL (>= 2)

Encoding UTF-8

RoxygenNote 7.2.3

NeedsCompilation no

Author Gianmarco Alberti [aut, cre]

Maintainer Gianmarco Alberti <gianmarcoalberti@gmail.com>

Repository CRAN

Date/Publication 2023-06-15 17:40:06 UTC

R topics documented:

---

caresid	*Correspondence Analysis plot with visualization of significant associations based on chi-square standardized residuals*

Description

Performs a Correspondence Analysis (CA) on a contingency table and creates a scatterplot of the row and column points on the selected dimensions. Optionally, the function can add segments to the plot to visualize significant associations between row and column categories on the basis of positive (unadjusted) standardized residuals larger than a given threshold. The segments can be optionally labelled with the corresponding residual value.

Visit this LINK to access the package's vignette.

Usage

```
caresid(
  cross.tab,
  dim1 = 1,
  dim2 = 2,
  segments = FALSE,
  category = NULL,
  mult.comp = FALSE,
  label.residuals = FALSE,
  residual.label.size = 2,
  dot.size = 1,
  dot.label.size = 2.5,
  axis.label.size = 9,
  square = FALSE
)
```

Arguments

cross.tab	A dataframe representing the input contingency table.
dim1	The first dimension to plot (default is 1).
dim2	The second dimension to plot (default is 2).
segments	Logical. If TRUE, add segments to the plot to connect row to column points (or viceversa) with positive standardized residuals larger than a given threshold (default is FALSE).
category	Character vector. If provided, only add segments from that/those row (or column) category(ies) to the column (or row) categories where the corresponding standardised residuals are positive and larger than a given threshold. If NULL (default) all the categories are considered.
mult.comp	Logical. If TRUE, adjust the residuals' significance threshold for multiple comparisons using Sidak's method (default is FALSE).
label.residuals	
	Logical. If TRUE, the value of the positive standardised residual will be shown as a label at the midpoint of every segment (default is FALSE).
residual.label.size	
	Numeric. The size of the residuals' label (default is 2).
dot.size	Numeric. The size of the scatterplot's points (default is 1).

`dot.label.size`	Numeric. The size of the points' label (default is 2.5).
`axis.label.size`	
	Numeric. The size of the axis labels (default is 9).
`square`	Logical. If TRUE, set the ratio of y to x to 1 (default is FALSE).

Details

If the `segment` argument is `FALSE` (default), a regular symmetric CA biplot is rendered.

If the `segment` argument is `TRUE`, the function adds segments to the plot to connect row and column points with positive (unadjusted) standardized residuals larger than a given threshold, indicating a significant association. The threshold is 1.96 if `mult.comp` is `FALSE`, and is adjusted for multiple comparisons if `mult.comp` is `TRUE`.

In the latter case, the threshold for significant residuals is calculated using the Sidak's method. It is based on an adjusted 0.05 alpha level which is calculated as `1-(1 - 0.05)^(1/(nr*nc))`, where `nr` and `nc` are the number of rows and columns in the table respectively. The adjusted alpha is then converted to a critical two-tailed z value (see Beasley-Schumacker 1995).

Please note, all the visualised associations (if any) are significant at least at alpha 0.05.

Optionally, the residual segments can be labelled with the corresponding residual value by setting the `label.residuals` to `TRUE`.

The idea of connecting points in a CA plot based on the value of standardized residuals can serve to visually highlight certain associations in your data. However, please note that while this function can help visualize the associations in the contingency table, it does not replace other formal approaches for the interpretation of the CA scatterplot and formal statistical tests for assessing the significance and strength of the association.

Value

A list with two elements:

- `stand.residuals` contains the unadjusted standardized residuals for all cells.
- `resid.sign.thres` contains the threshold used to determine significant residuals.

References

Beasley TM and Schumacker RE (1995), Multiple Regression Approach to Analyzing Contingency Tables: Post Hoc and Planned Comparison Procedures, The Journal of Experimental Education, 64(1): 86, 89.

Examples

```
# Create a toy dataset (famous Eye-color Hair-color dataset)

mytable <- structure(list(BLACK_H = c(68, 20, 15, 5),
BROWN_H = c(119, 84, 54, 29),
```

```
RED_H = c(26, 17, 14, 14),
BLOND_H = c(7, 94, 10, 16)),
class = "data.frame",
row.names = c("Brown_E", "Blue_E", "Hazel_E", "Green_E"))

# EXAMPLE 1
# Run the function:

result <- caresid(mytable, segments=TRUE)


# EXAMPLE 2
# As above, but adjusting for multiple comparisons:

result <- caresid(mytable, segments=TRUE, mult.comp=TRUE)


# EXAMPLE 3
# As in the first example, but selecting only 2 row categories;
# residual labels are shown:

result <- caresid(mytable, segments=TRUE, category=c("Brown_E", "Green_E"), label.residuals=TRUE)
```

Index