Package 'Path.Analysis'

January 20, 2025

Type Package

Title Path Coefficient Analysis

Version 0.1

Date 2024-9-12.

Author Ali Arminian [aut, cre, cph] (<https://orcid.org/0000-0003-4749-6085>)

Maintainer Ali Arminian <abeyran@gmail.com>

Description Facilitates the performance of several analyses, including simple and sequential path coefficient analysis, correlation estimate, drawing correlogram, Heatmap, and path diagram. When working with raw data, that includes one or more dependent variables along with one or more independent variables are available, the path coefficient analysis can be conducted. It allows for testing direct effects, which can be a vital indicator in path coefficient analysis. The process of preparing the dataset rule is explained in detail in the vignette file ``Path.Analysis_manual.Rmd''. You can find this in the folders labelled ``data'' and ``~/inst/extdata''. Also see: 1)the 'lavaan', 2)a sample of sequential path analysis in 'metan' suggested by Olivoto and Lúcio (2020) <doi:10.1111/2041-210X.13384>, 3)the simple 'PATHSAS' macro written in 'SAS' by Cramer et al. (1999) <doi:10.1093/jhered/90.1.260>, and 4)the semPlot() function of 'OpenMx' as initial tools for conducting path coefficient analyses and SEM (Structural Equation Modeling). To gain a comprehensive understanding of path coefficient analysis, both in theory and practice, see a 'Minitab' macro developed by Arminian, A. in the paper by Arminian et al. (2008) <doi:10.1080/15427520802043182>.

License GPL-3

URL https://github.com/abeyran/Path.Analysis

BugReports https://github.com/abeyran/Path.Analysis/issues

Depends R (>= 4.1.0)

Imports stats, corrr, corrplot, Hmisc, gplots, mathjaxr, pastecs, graphics, grDevices, DiagrammeR, ComplexHeatmap, metan

Suggests car, ggplot2, devtools, usethis, testthat, knitr, rmarkdown, roxygen2, spelling

VignetteBuilder knitr

RdMacros mathjaxr

Encoding UTF-8
Copyright Ali Arminian
RoxygenNote 7.3.2
Language en-US
LazyData true
LazyLoad true
NeedsCompilation no
BuildManual TRUE
Repository CRAN
Date/Publication 2024-09-25 08:20:05 UTC

Contents

`corr` *Correlation Analysis*

Description

[Stable]

- `corr()` estimates Pearson correlation coefficients among parametric numerical characteristics as follows:
- `The Pearson correlation coefficient:`

$$r_{x,y} = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}}$$

or:

$$r_{x,y} = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\sqrt{\Sigma(x - \bar{x})^2 \Sigma(y - \bar{y})^2}}$$

where $r_{x,y}$ is the `correlation coefficient` between x and y variables.

Usage

```
corr(datap, verbose = FALSE)
```

Arguments

`datap`	The data set
`verbose`	If `verbose = TRUE` then some results are printed in the console.

Details

The `corr()` function estimates correlation coefficients and their significance in the form of a table of one or more independent (exogenous) variables on a dependent (endogenous) variable along with testing the significance.

Value

Returns a list of two objects:

Correlations the data frame of Pearson's correlation coefficients

P_values the data frame of significance of correlation coefficients (r):

- `p` p-value for testing the r
- `lowCI` lower confidence interval of r
- `uppCI` upper confidence interval of r

Author(s)

Ali Arminian abeyran@gmail.com

See Also

`correlation`

Examples

```
data(dtsimp)
corr(dtsimp, verbose = FALSE)


data(dtraw)
corr(dtraw[, -1], verbose = FALSE)
```

`cor_plot` *Drawing the correlogram*

Description

[Stable]

- `corr_plot()` draws a correlogram for data

Usage

```
cor_plot(datap)
```

Arguments

`datap`	The data set

Value

Returns an object of class `gg, ggmatrix`.

Author(s)

Ali Arminian abeyran@gmail.com

References

Olivoto, T, and A Dal'Col Lúcio. 2020. "Metan: An r Package for Multi-environment Trial Analysis." Methods in Ecology and Evolution, 11(6): 783–89. https://doi.org/10.1111/2041-210 X.13384.

See Also

`correlogram`, `diagram`, and `lavaan` package for drawing path diagrams.

Examples

```
data(dtsimp)
cor_plot(dtsimp)
```

dataprep *Data preparation*

Description

[Experimental]

Prepares data for analyses

Usage

```
dataprep(datap)
```

Arguments

datap	dataset

Value

Returns a data frame

desc *Descriptive statistics*

Description

[Experimental]

- `desc()` estimates the descriptive statistics such as `Min`(Minimum), `1st Qu.`(quartile), `Median`, `Mean` (average), `3rd Qu.`(3rd quartile), `Max`(maximum), `var` (variance), `std.dev`(standard deviation), `coef.var` (CV or coefficient of variation) of the data set.

Usage

```
desc(datap, resp)
```

Arguments

datap	The data set
resp	an integer value indicating the column in `datap` that corresponds to the response variable.

Details

The `desc()` function estimates the descriptive statistics, in tables for one or more independent (exogenous) variables on a dependent(endogenous) variable. It acts only on numerical variables. For example for the variable `x`:

- `1st. quartile`:
$$Q_1 = (n+1)x1/4$$
- `2nd. quartile or Median`:
$$md = (n+1)x2/4$$
- `3rd Qu.`:
$$Q_3 = (n+1)x3/4$$
- `Arithmetic mean`:
$$\bar{x} = \frac{1}{n}\sum_{i=i}^{n} x_i$$
- `Range`:
$$R_x = \max(x) - \min(x)$$
- `Variance`:
$$\sigma_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$$
- `Standard deviation`:
$$sd_x = \sqrt{\frac{\sum_i (x_i - \mu)^2}{n}}$$
- `SEM or SE.mean`, the standard error of the mean is calculated simply by taking the standard deviation and dividing it by the square root of the sample size:
$$SEM_x = \frac{sd(x)}{\sqrt{n}}$$
- `coef.var or coefficient of variation`:
$$CV = \frac{sd(x)}{\bar{x}} \times 100$$

Value

Returns a list of 3 objects:

desc1 Descriptive statistics1 of input data

desc2 Descriptive statistics2 of input data

corcf A table of correlation coefficients

Author(s)

Ali Arminian abeyran@gmail.com

References

Bhattacharyya GK and Johnson RA 1997. Statistical Concepts and Methods, John Wiley and Sons, New York.

Draper N and Smith H 1981. Applied Regression Analysis, John Wiley & Sons, New York.

Neter, J, Whitmore, GA, Wasserman, W 1992. Applied Statistics. Allyn & Bacon, Incorporated, ISBN 10: 0205134785 / ISBN 13: 9780205134786.

Snedecor, G.W., Cochran, W.G. 1980. Statistical Methods. Iowa State University Press.

See Also

`correlation`, `multiple linear regression`,

Examples

```
data(dtsimp)
desc(dtsimp, 1)


data(dtraw)
desc(dtraw[, -1], 1)

data(heart)
desc(heart, 2)
```

`dtraw` *Dataset 2: a number of 9 traits measured on 35 Camelina DH lines.*

Description

Dataset 2: a number of 9 traits measured on 35 Camelina DH lines.

Usage

```
data(dtraw)
```

Format

A `data.frame` with 35 observations of 9 variables.

`DH lines` a character vector

`y` a numeric vector

`X1` a numeric vector

`X2` a numeric vector

`X3` a numeric vector

`X4` a numeric vector

`X5` a numeric vector

`X6` a numeric vector

`X7` a numeric vector

`X8` a numeric vector

Examples

```
library(Path.Analysis)
data(dtraw)
```

`dtraw2` *Dataset 3: a number of 9 traits measured on 35 Camelina DH lines.*

Description

Dataset 3: a number of 9 traits measured on 35 Camelina DH lines.

Usage

```
data(dtraw2)
```

Format

A `data.frame` with 35 observations of 9 variables.

`DH lines` a character vector considered as rownames

`y` a numeric vector

`X1` a numeric vector

`X2` a numeric vector

`X3` a numeric vector

`X4` a numeric vector

`X5` a numeric vector

`X6` a numeric vector

`X7` a numeric vector

`X8` a numeric vector

Examples

```
library(Path.Analysis)
data(dtraw2)
```

dtseq *Dataset 4: a dataframe consisting of 7 variables measured on 8 observations.*

Description

Dataset 4: a dataframe consisting of 7 variables measured on 8 observations.

Usage

```
data(dtseq)
```

Format

A `data.frame` with 8 observations of 7 variables.

`Genotypes` a character vector

`YLD` a numeric vector

`DFT` a numeric vector

`FS` a numeric vector

`FV` a numeric vector

`FW` a numeric vector

`DFL` a numeric vector

`FLP` a numeric vector

Examples

```
library(Path.Analysis)
data(dtseq)
```

dtseqr *Dataset5*

Description

Dataset5

Usage

```
data(dtseqr)
```

Format

A `data.frame` with 24 observations of 7 variables.

`Genotypes` a character vector

`Rep` a numeric vector

`YLD` a numeric vector

`DFT` a numeric vector

`FS` a numeric vector

`FV` a numeric vector

`FW` a numeric vector

`DFL` a numeric vector

`FLP` a numeric vector

Examples

```
library(Path.Analysis)
data(dtseqr)
```

`dtsimp` *Dataset 1: a dependent (y) and 3 independent(x1 to x3) variables.*

Description

Dataset 1: a dependent (y) and 3 independent(x1 to x3) variables.

Usage

```
data(dtsimp)
```

Format

A `data.frame` with 105 observations of 4 variables.

`y` a numeric vector

`x1` a numeric vector

`x2` a numeric vector

`x3` a numeric vector

Examples

```
library(Path.Analysis)
data(dtsimp)
```

heart *Dataset 6: Heart Disease data set*

Description

A mixed variable dataset containing 14 variables of 297 patients for their heart disease diagnosis.

Usage

```
data(heart)
```

Format

A `data.frame` including 297 rows and 14 variables:

age Age in years (numerical).

sex Sex: 1 = male, 0 = female (logical).

`heart.disease` a numeric vector as dependent.

`biking` a numeric vector as the first independent.

`smoking` a numeric vector as the 2nd independent.

Source

The data set is belong to machine learning repository of UCI. The original data set includes 303 patients with 6 NA's. After removing missing values, it reduced into 297 patients.

https://archive.ics.uci.edu/ml/datasets/Heart+Disease

References

Lichman, M. (2013). UCI machine learning repository.

Examples

```
library(Path.Analysis)
data(heart)
```

heat_map *Creating the* Heatmap *chart*

Description

[Stable]

- heat_map() draws a double-clustered heatmap for path coefficients analysis. Please be cautious that this function acts only on numeric variables/columns (see example on dtraw2 data set). Users for drawing other types of heatmaps may use heatmap.3, ComplexHeatmap and pheatmap R packages. Where an example is given in the vignette manual of this package (Path.Analysis_manual.Rmd)

Usage

```
heat_map(datap)
```

Arguments

datap	The data set

Value

Returns an object of class heatmap.2.

Author(s)

Ali Arminian abeyran@gmail.com

See Also

lavaan and diagram packages for drawing path diagrams.

Examples

```
data(dtraw2)
dtraw2 <- scale(as.data.frame(dtraw2))
heat_map(dtraw2)
```

matdiag *Direct and Indirect Effects Matrices and Diagram*

Description

[Stable]

- matdiag() extracts the direct effect and indirect effects matrices of data in path analysis along with the significance of direct effects where direct effects are shown as a vector (columnar matrix of 1*n dimensions and indirect effects are off-diagonal effects. Later, draws a diagram for path coefficient analysis based on the DiagrammeR package.

Usage

```
matdiag(datap, resp, verbose = FALSE)
```

Arguments

datap	The data set
resp	The response variable
verbose	If verbose = TRUE then some results are printed

Details

The matdiag function estimates the direct and indirect effects in path coefficient analysis as tables along with drawing the diagram of path analysis. This is apparently the only program testing the significance of direct effects in a path analysis. Note: all variables must be numeric for matrix calculations and the next plotting.

- In a path model, path coefficients or direct effects (Pi's) indicate the direct effects of a variable on another, and are standardized partial regression coefficients (in Wright's terminology) due they are estimated from correlations or from the transformed (standardized) data as:

$$P_i = \beta_i \frac{\sigma_{X_i}}{\sigma_Y}$$

- The path equations are as follows:
- One dependent variable:

$$P_1 + P_2 r_{12} + P_3 r_{13} + ... + P_n r_{1n} = rY_1$$

$$P_1 r_{21} + P_2 + P_3 r_{23} + ... + P_n r_{2n} = rY_2$$

$$...$$

$$P_1 rn_1 + P_2 r_{n2} + P_3 r_{n3} + ... + P_n = rY_n$$

- Extension to more dependent variables: `Path.Analysis` is capable of performing this straightforward function through detailed explanations. The linear regression model with a single response in its form is as follows (Johnson and Wichern (2007): $Y = \beta_0 + \beta_1 Z_1 + ... + \beta_r Z_r + \epsilon$ where the multivariate multiple linear regression model is as follows:

$$Y_1 = \beta_0 + \beta_1 Z_{11} + \beta_2 Z12 + ... + \beta_r Z_{1r} + \epsilon_1$$

$$Y_2 = \beta_0 + \beta_1 Z_{21} + \beta_2 Z22 + ... + \beta_r Z_{2r} + \epsilon_2$$

$$...$$

$$Y_n = \beta_0 + \beta_1 Z_{n1} + \beta_2 Zn2 + ... + \beta_r Z_{nr} + \epsilon_n$$

As stated by Bondari (1990), for two dependent variables Y_1 and Y_2:

$$Y_1 = p_1 X_1 + p_2 X_2 + p_3 X_3 + ... + p_n X_n$$

$$Y_2 = p'_1 X_1 + p'_2 X_2 + p'_3 X_3 + ... + p'_n X_n$$

$$...$$

where:

$$r_{Y_1 Y_2} = p_1 p'_1 + p_2 p'_2 + p_3 p'_3 + ... + p_n p'_n + \sigma_{i=j} p_i p'_1 r_{ij} = \sigma_{i,j} p_i p'_i r_{ij}$$

Value

Returns a list with three objects

direff a data frame of direct effects

matall a matrix of direct and indirect effects

Residual a constant of residuals

Author(s)

Ali Arminian abeyran@gmail.com

References

Arminian, A, MS Kang, M Kozak, S Houshmand, and P Mathews. 2008. “MULTPATH: A Comprehensive Minitab Program for Computing Path Coefficients and Multiple Regression for Multivariate Analyses.” Journal of Crop Improvement, 22(1): 82–120.

Bondari, K. 1990. "PATH ANALYSIS IN AGRICULTURAL RESEARCH," Conference on Applied Statistics in Agriculture. https://do i.org/10.4148/2475-7772.1439

Cramer, C.S, TC Wehner, and SB Donaghy. 1999. “PATHSAS: A SAS Computer Program for Path Coefficient Analysis of Quantitative Data.” Journal of Heredity, 90(1): 260–62. https://doi.org/10 .1093/jhered/90.1.260.

Johnson, R.A., Wichern, D.W. 2007. Applied Multivariate Statistical Analysis. Prentice Hall, USA.

Li, C.C. 1975. Path Analysis: A Primer. Boxwood Pr. 346 p.

Olivoto, T, and A Dal’Col Lúcio. 2020. “Metan: An r Package for Multi-environment Trial Analysis.” Methods in Ecology and Evolution, 11(6): 783–89. https://doi.org/10.1111/2041-210 X.13384.

Wolfle, LM. 2003. "The Introduction of Path Analysis to the Social Sciences, and Some Emergent Themes: An Annotated Bibliography." Structural Equation Modeling, 10(1): 1–34.

Wright, S. 1923. "The Theory of Path Coefficients a Reply to Niles's Criticism." Genetics, 8(3): 239.

———. 1934. "The Method of Path Coefficients." The Annals of Mathematical Statistics, 5(3): 161–215.

———. 1960. "Path Coefficients and Path Regressions: Alternative or Complementary Concepts?" Biometrics, 16(2): 189–202.

See Also

`correlation`, `multiple linear regression`, and matrix notations in mathematics.

`lavaan` and `diagrammeR` packages for drawing path diagrams

Examples

```
data(dtsimp)
matdiag(dtsimp, 1, verbose = FALSE)

data(dtraw)
matdiag(dtraw[, -1], 1, verbose = FALSE)

data(heart)
matdiag(heart, 2, verbose = FALSE)
```

`network.plot` *Network plot*

Description

[Stable]

- `network.plot()` draws the network plot of path coefficients analysis

Usage

```
network.plot(datap)
```

Arguments

`datap`	The data set

Details

The `network.plot()` draws a correlogram and a heatmap for data, if requested by user

Value

Returns an object of class `network_plot`.

Author(s)

Ali Arminian abeyran@gmail.com

References

Kuhn et al. 2022. corrr package. doi: <10.32614/CRAN.package.corrr> https://github.com/tidymodels/corrr

See Also

`correlogram`, `diagram`, and `lavaan` packages for drawing path diagrams.

Examples

```
data(dtraw2)
network.plot(dtraw2)
```

`Path.Analysis` *Path Coefficient Analysis*

Description

Path.Analysis does descriptive statistics on dataset and importantly graphical representation of data such as drawing heatmaps, correlogram and path diagram.

Author(s)

Ali Arminian abeyran@gmail.com

See Also

Useful links:

- https://github.com/abeyran/Path.Analysis
- Report bugs at https://github.com/abeyran/Path.Analysis/issues

reg *Multiple Linear Regression*

Description

[Experimental]

- `reg()` performs a multiple linear regression analysis with extracting the attributed parameters

Usage

```
reg(datap, resp, verbose = FALSE)
```

Arguments

`datap`	The data set
`resp`	an integer value indicating the column in `datap` that
`verbose`	If `verbose = TRUE` then some results are printed in the console. corresponds to the response variable.

Details

The `reg` function fits a multiple linear regression analysis of one or more independent (exogenous) variables on a dependent(endogenous) variable in a linear pattern along with testing the significance of parameters. It is important that according to the type of data may produce some warning errors e.g., for dtsimp as: Warning message: In summary.lm(mlreg): essentially perfect fit: summary may be unreliable. This case is due to the intrinsic characteristics of data

Value

An object of class list

Author(s)

Ali Arminian abeyran@gmail.com

See Also

`multiple linear regression`

Examples

```
data(dtsimp)
reg(dtsimp, 1, verbose = FALSE)


data(heart)
reg(heart, 1, verbose = FALSE)
```

Index