Package 'MCCM'

April 18, 2024

Title Mixed Correlation Coefficient Matrix

Version 0.1.0

Description The IRLS (Iteratively Reweighted Least Squares) and GMM (Generalized Method of Moments) methods are applied to estimate mixed correlation coefficient matrix (Pearson, Polyseries, Polychoric), which can be estimated in pairs or simultaneously. For more information see Peng Zhang and Ben Liu (2024) <doi:10.1080/10618600.2023.2257251>; Ben Liu and Peng Zhang (2024) <doi:10.

License GPL

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

Suggests knitr, rmarkdown

Depends R (>= 3.5.0),

Imports MASS, polycor, lavaan, mvtnorm

NeedsCompilation no

Author Ben Liu [aut, cre],
Peng Zhang [ths],
Xiaowei Lou [aut, dtc]

Maintainer Ben Liu <12035024@zju.edu.cn>

Repository CRAN

Date/Publication 2024-04-18 14:02:52 UTC

R topics documented:

`CECERS` *Chinese Early Childhood Environment Rating Scale*

Description

The CECERS uses a 9-point scoring system, 1-3 (inadequate), 5 (least acceptable), 7 (good), and 9 (excellent), to measure the quality of Chinese early children education (ECE) programs for children aged 3 to 6. The CECERS has a total of 51 items organized in eight categories: (1) Space and Furnishings (9 items); (2) Personal Care Routines (6 items); (3) Curriculum Planning and Implementation (5 items); (4) Whole-Group Instruction (7 items); (5) Activities (9 items); (6) Language-Reasoning (4 items); (7) Guidance and Interaction (5 items); (8) Parents and Staff (6 items).

Format

A data frame with 1383 rows and 95 variables:

Source

Kejian Li, Peng Zhang, Bi Ying Hu, Margaret R Burchinal, Xitao Fan, and Jinliang Qin. Testing the 'thresholds' of preschool education quality on child outcomes in china. Early Childhood Research Quarterly, 47:445–456, 2019.

`dphixy` *Scaled Bivariate Normal Density*

Description

Bivariate normal density with mean 0 variance 1.

Usage

```
dphixy(x, y, rho)
```

Arguments

`x, y`	points value.
`rho`	correlation coefficient.

Value

the density value.

Examples

```
library(mvtnorm)
dmvnorm(c(1,-1),sigma = matrix(c(1,0.5,0.5,1),2,2))
dphixy(1,-1,0.5)
```

---

draw_correlation_matrix   *Draw the Correlation Matrix*

Description

Estimate the MCCM from dataframe and draw it with scatter plot of matrices (SPLOM). With bivariate scatter plots below the diagonal, histograms on the diagonal, and the polychoric correlation coefficients with standard errors above the diagonal. Correlation ellipses are drawn in the same graph. The red lines below the diagonal are the LOESS smoothed lines, fitting a smooth curve between two variables.

Usage

```
draw_correlation_matrix(
  data1,
  order_indx,
  pair_est = FALSE,
  MLE = FALSE,
  R0 = NULL,
  app = TRUE,
  korder = 2,
  max_iter = 1000,
  max_tol = 1e-08,
  show_log = FALSE
)
```

Arguments

`data1`	a dataframe containing continuous or ordinal variable.
`order_indx`	a vector to indicate the ordinal variables.
`pair_est`	bool value, TRUE for pairwise estimation, FALSE for simultaneous estimation.

MLE	bool value, TRUE for maximum likelihood estimation, FALSE for IRLS (pairwise) or IGMM (simultaneous) estimation.
R0	the initial value for correlation vector, default Pearson correlation matrix.
app	bool value for approximation, TRUE for Legendre approximation, FALSE for common integral.
korder	the order of Legendre approximation.
max_iter	max iteration number for IGMM.
max_tol	max tolerance for iteration algorithm.
show_log	bool value, TRUE for showing calculation log.

Value

the SPLOM plot.

See Also

MCCM_est, summary_MCCM_est

Examples

```
library(mvtnorm)
library(MASS)
library(polycor)
library(lavaan)
set.seed(1997)
n = 10000
rho12=0.3
rho13=0.4
rho14=0.5
rho23=0.6
rho24=0.7
rho34=0.8

R = matrix(c(1,rho12,rho13,rho14,rho12,1,rho23,rho24,rho13,rho23,1,rho34,
rho14,rho24,rho34,1),4,4)
indc = c(3,4)
thresholds = list(c(),c(),0,0)
data1 = gen_mixed(n=n,R=R,indc=indc,thresholds=thresholds)
data2 = data.frame(data1$observed)

# pairwise MLE estimation
draw_correlation_matrix(data2,indc,TRUE,TRUE)
# pairwise IRLS estimation
draw_correlation_matrix(data2,indc,TRUE,FALSE)
# simultaneous MLE estimation
draw_correlation_matrix(data2,indc,FALSE,TRUE)
# simultaneous IGMM estimation
draw_correlation_matrix(data2,indc,FALSE,FALSE)
```

esti_polychoric *Polychoric Correlation*

Description

Estimate the polychoric correlation coefficient.

Usage

```
esti_polychoric(X, maxn = 100, e = 1e-08, ct = FALSE)
```

Arguments

X	a matrix(2*N) or dataframe contains two polychoric variable, or a contingency table with both columns and rows names.
maxn	the maximum iterations times.
e	the maximum tolerance of convergence.
ct	TRUE for contingency table, FALSE for matrix or dataframe

Value

rho	estimated value of polychoric correlation coefficient.
std	standard deviation of rho.
iter	times of iteration convergence.
Ex,Ey	the support points series of regression model

References

Zhang, P., Liu, B., & Pan, J. (2024). Iteratively Reweighted Least Squares Method for Estimating Polyserial and Polychoric Correlation Coefficients. Journal of Computational and Graphical Statistics, 33(1), 316–328. https://doi.org/10.1080/10618600.2023.2257251

See Also

esti_polyserial

Examples

```
X = gen_polychoric(1000,0.5,0:1,-1:0)
result = esti_polychoric(X)
print(c(result$rho,result$std,result$iter))
```

`esti_polyserial` *Polyserial Correlation*

Description

Estimate the polyserial correlation coefficient.

Usage

```
esti_polyserial(X, maxn = 100, e = 1e-08)
```

Arguments

`X`	a matrix(2*N) or dataframe contains two polyserial variable(Continuous variable first).
`maxn`	the maximum iterations times.
`e`	the maximum tolerance of convergence.

Value

`rho`	estimated value of polyserial correlation coefficient.
`std`	standard deviation of rho.
`iter`	times of iteration convergence.
`Ex,Ey`	the support point of regression model.

References

Zhang, P., Liu, B., & Pan, J. (2024). Iteratively Reweighted Least Squares Method for Estimating Polyserial and Polychoric Correlation Coefficients. Journal of Computational and Graphical Statistics, 33(1), 316–328. https://doi.org/10.1080/10618600.2023.2257251

See Also

esti_polychoric

Examples

```
X = gen_polyseries(1000,0.5,-1:1)
result = esti_polyserial(X)
result
```

est_mixedGMM *Estimating Mixed Correlation Matrix by IGMM*

Description

An accelerated function to estimate a mixed correlation coefficient matrix, as well as its covariance matrix, for dataframes containing continuous and ordinal variable.

Usage

```
est_mixedGMM(
  dataYX,
  order_indx,
  R0 = NULL,
  app = TRUE,
  korder = 2,
  max_iter = 1000,
  max_tol = 1e-08,
  show_log = FALSE
)
```

Arguments

dataYX	a dataframe or matrix containing both continuous and ordinal variables.
order_indx	a vector to indicate the ordinal variables.
R0	the initial value for correlation vector, default Pearson correlation matrix.
app	bool value for approximation, TRUE for Legendre approximation, FALSE for common integral.
korder	the order of Legendre approximation.
max_iter	max iteration number for IGMM.
max_tol	max tolerance for iteration algorithm.
show_log	bool value, TRUE for showing calculation log.

Value

Rhat	The estimated correlation coefficients.
COV	The estimated covariance matrix for Rhat

References

arXiv:2404.06781

Examples

```
library(mvtnorm)
library(MASS)
set.seed(1997)
n = 500
rho12=0.3
rho13=0.4
rho14=0.5
rho23=0.6
rho24=0.7
rho34=0.8

R = matrix(c(1,rho12,rho13,rho14,rho12,1,rho23,rho24,rho13,rho23,1,rho34,
rho14,rho24,rho34,1),4,4)
indc = c(3,4)
thresholds = list(c(),c(),0,0)
data1 = gen_mixed(n=n,R=R,indc=indc,thresholds=thresholds)
data2 = data.frame(data1$observed)
out1 = est_mixedGMM(dataYX = data2,order_indx = indc)
print(out1$Rhat)
print(out1$COV)
```

---

est_thre *Thresholds Estimation*

Description

Function to calculate thresholds from ordinal variables.

Usage

```
est_thre(X)
```

Arguments

X	a ordinal series.

Value

the estimated value for thresholds.

Examples

```
library(mvtnorm)
set.seed(1997)
R1 = gen_CCM(4)
n = 1000
indc = 3:4
thresholds = list(c(),c(),c(-1),c(1))
```

```
data1 = gen_mixed(n,R1,indc,thresholds=thresholds)$observed
est_thre(data1[,3])
est_thre(data1[,4])
```

---

gen_CCM *Positive Semidefinite Correlation Matrix*

Description

Generate a positive semidefinite correlation coefficients matrix

Usage

```
gen_CCM(d)
```

Arguments

d	the dimension of matrix.

Value

a correlation coefficients matrix.

Examples

```
X = gen_CCM(4)
print(X)
```

---

gen_mixed *Continuous and Ordinal Simulated Data*

Description

Generate multi-normal sample and segment it into ordinal.

Usage

```
gen_mixed(n, R, indc, thresholds)
```

Arguments

n	the sample size.
R	the correlation coefficient matrix.
indc	vector to indicate whether variables are continuous or categorical.
thresholds	list contains thresholds for ordinal variables

Value

`latent`	the original normal data.
`observed`	the observed ordinal data.

Examples

```
library(mvtnorm)
set.seed(1997)
R1 = gen_CCM(6)
n = 1000
indc = 4:6
thresholds = list(
  c(),
  c(),
  c(),
  c(0),
  c(-0.5,0),
  c(0,0.5)
)
data1 = gen_mixed(n,R1,indc,thresholds)$observed
data1 = data.frame(data1)
table(data1$X4,data1$X5)
table(data1$X5,data1$X6)
```

`gen_polychoric` *Generate Polychoric Sample*

Description

Generate polychoric sample with hidden distribution: binormal with correlation coefficient rho.

Usage

```
gen_polychoric(n, rho, a, b)
```

Arguments

`n`	sample size.
`rho`	correlation coefficient.
`a`	the cutoff points array.
`b`	the cutoff points array.

Value

Polychoric sample with size n(in a 2*n matrix).

See Also

gen_polyseries gen_rho

Examples

```
gen_polychoric(100,0.5,-1:1,1:2)
```

`gen_polyseries` *Generate Polyseries Sample*

Description

Generate polyseries sample with hidden distribution: binormal with correlation coefficient rho.

Usage

```
gen_polyseries(n, rho, a)
```

Arguments

n	sample size.
rho	correlation coefficient.
a	the cutoff points array.

Value

Polyseries sample with size n(in a 2*n matrix).

See Also

gen_rho gen_polychoric

Examples

```
gen_polyseries(100,0.5,-1:1)
```

gen_rho *Generate Specific Binormal Distribution*

Description

Generate random number of binormal distribution with 0 mean unit variance and correlation coefficient rho.

Usage

```
gen_rho(n, rho)
```

Arguments

`n`	sample size.
`rho`	correlation coefficient.

Value

Binormal random number with length n(in a 2*n matrix).

See Also

gen_polyseries gen_polychoric

Examples

```
gen_rho(100,0.5)
```

mb *Mean Bias*

Description

Calculate the MB of an array of estimates relative to the true value.

Usage

```
mb(rhohat, rho)
```

Arguments

`rhohat`	an array of estimators of rho.
`rho`	the true value of rho.

Value

the mean bias of rhohat array.

See Also

mrb rmse

Examples

```
rho = 0.5
rhohat = 0.5 + rnorm(10)
mb(rhohat,rho)
```

MCCM_est — *General Function to Estimate Mixed Correlation Coefficient Matrix*

Description

Estimate the correlation matrix for dataframes containing continuous and ordinal variable, in pairs or simultaneously, using MLE, IRLS, or IGMM.

Usage

```
MCCM_est(
  dataYX,
  order_indx,
  pair_est = FALSE,
  MLE = FALSE,
  R0 = NULL,
  app = TRUE,
  korder = 2,
  max_iter = 1000,
  max_tol = 1e-08,
  show_log = FALSE
)
```

Arguments

dataYX	a dataframe or matrix containing both continuous and ordinal variables.
order_indx	a vector to indicate the ordinal variables.
pair_est	bool value, TRUE for pairwise estimation, FALSE for simultaneous estimation.
MLE	bool value, TRUE for maximum likelihood estimation, FALSE for IRLS (pairwise) or IGMM (simultaneous) estimation.
R0	the initial value for correlation vector, default Pearson correlation matrix.
app	bool value for approximation, TRUE for Legendre approximation, FALSE for common integral.

korder	the order of Legendre approximation.
max_iter	max iteration number for IGMM.
max_tol	max tolerance for iteration algorithm.
show_log	bool value, TRUE for showing calculation log.

Value

Rmatrix	Estimated mixed correlation coefficient matrix.
std_matrix	Estimated standard deviation for each mixed correlation coefficient.
COV	The covariance matrix for MCCM (simultaneous estimation only).

See Also

esti_polyserial, esti_polychoric, est_mixedGMM, summary_MCCM_est, draw_correlation_matrix

Examples

```
library(mvtnorm)
library(MASS)
library(polycor)
library(lavaan)
set.seed(1997)
n = 10000
rho12=0.3
rho13=0.4
rho14=0.5
rho23=0.6
rho24=0.7
rho34=0.8

R = matrix(c(1,rho12,rho13,rho14,rho12,1,rho23,rho24,rho13,rho23,1,rho34,
rho14,rho24,rho34,1),4,4)
indc = c(3,4)
thresholds = list(c(),c(),0,0)
data1 = gen_mixed(n=n,R=R,indc=indc,thresholds=thresholds)
data2 = data.frame(data1$observed)

# pairwise MLE estimation
out_pair_MLE = MCCM_est(dataYX=data2,order_indx=indc,pair_est=TRUE,MLE=TRUE)
# pairwise IRLS estimation
out_pair_IRLS = MCCM_est(dataYX=data2,order_indx=indc,pair_est=TRUE,MLE=FALSE)
# simultaneous MLE estimation
out_sim_MLE = MCCM_est(dataYX=data2,order_indx=indc,pair_est=FALSE,MLE=TRUE)
# simultaneous IGMM estimation
out_sim_IGMM = MCCM_est(dataYX=data2,order_indx=indc,pair_est=FALSE,MLE=FALSE)

summary_MCCM_est(out_pair_MLE)
summary_MCCM_est(out_pair_IRLS)
summary_MCCM_est(out_sim_MLE)
summary_MCCM_est(out_sim_IGMM)
```

`mrb` *Mean Relative Bias*

Description

Calculate the MRB of an array of estimates relative to the true value.

Usage

```
mrb(rhohat, rho)
```

Arguments

`rhohat`	an array of estimators of rho.
`rho`	the true value of rho.

Value

the mean relative bias of rhohat array.

See Also

mb rmse

Examples

```
rho = 0.5
rhohat = 0.5 + rnorm(10)
mrb(rhohat,rho)
```

`Parenteral_nutrition` *Parenteral_nutrition*

Description

The Parenteral Nutrition data were collected from 543 patients of whom 386 were given parenteral nutrition alone, 145 were given enteral and parenteral nutrition, and 3 were given enteral nutrition only. There are 23 main discrete variables, such as: clinical stages(1-4), dietary status(1-3), NRS(0-6), PG-SGA-qualitative(1-3), etc.

Format

A data frame with 1086 rows and 29 variables:

Phixy — *Scaled Bivariate Normal Approximation*

Description

Standard bivariate normal distribution approximated with Legendre polynomials.

Usage

```
Phixy(x, y, rho, korder = 3, app = TRUE)
```

Arguments

`x, y`	P(X<=x,Y<=y).
`rho`	correlation coefficient.
`korder`	order of Legendre approximation.
`app`	bool value TRUE for approximation, FALSE for integral.

Value

P(X<=x,Y<=y).

Examples

```
library(mvtnorm)
pmvnorm(upper = c(1,-1),sigma = matrix(c(1,0.5,0.5,1),2,2))
Phixy(1,-1,0.5,2,app=TRUE)
Phixy(1,-1,0.5,app=TRUE)
```

rmse — *Root Mean Squared Error*

Description

Calculate the RMSE of an array of estimates relative to the true value.

Usage

```
rmse(rhohat, rho)
```

Arguments

`rhohat`	an array of estimators of rho.
`rho`	the true value of rho.

Value

the root mean squared error of rhohat array.

See Also

mb mrb

Examples

```
rho = 0.5
rhohat = 0.5 + rnorm(10)
rmse(rhohat,rho)
```

---

summary_MCCM_est *Summary a MCCM Estimation Result*

Description

Display the estimated correlation matrix and std matrix for a MCCM_est list.

Usage

```
summary_MCCM_est(out_MCCM)
```

Arguments

`out_MCCM`	output of function MCCM_est.

Value

The summary of estimation.

See Also

MCCM_est, draw_correlation_matrix

Index