Package: mult.latent.reg (via r-universe)

October 25, 2024

Type Package

Title Regression and Clustering in Multivariate Response Scenarios

Version 0.2.0

Description Fitting multivariate response models with random effects
on one or two levels; whereby the (one-dimensional) random
effect represents a latent variable approximating the
multivariate space of outcomes, after possible adjustment for
covariates. The method is particularly useful for multivariate,
highly correlated outcome variables with unobserved
heterogeneities. Applications include regression with
multivariate responses, as well as multivariate clustering or
ranking problems. See Zhang and Einbeck (2024)
<doi:10.1007/s42519-023-00357-0>.

License GPL-3

Imports mvtnorm, stats, matrixStats, utils, lme4

Encoding UTF-8

LazyData true

RoxygenNote 7.3.2

Depends R (>= 3.5.0)

Collate 'fetal_covid_data.R' 'IALS_data.R' 'mult.latent.reg-package.R'
'mult.reg_1level.R' 'mult.reg_2level.R' 'start.em.1level.R'
'start.em.2level.R' 'start.em.R' 'trading_data.R'
'twins_data.R'

Config/testthat/edition 3

NeedsCompilation no

Author Yingjuan Zhang [aut, cre], Jochen Einbeck [aut, ctb]

Maintainer Yingjuan Zhang <yingjuan.zhang@durham.ac.uk>

Repository CRAN

Date/Publication 2024-10-24 19:00:02 UTC

Contents

---

fetal_covid_data	*A set of fetal movements data collected before and during the Covid-19 pandemic*

Description

The data were recorded via 4D ultrasound scans from 40 fetuses (20 before Covid and 20 during Covid) at 32 weeks gestation, and consist of the number of movements each fetus carries out in relation to the recordable scan length.

Usage

```
data(fetal_covid_data)
```

Format

An object of class `"data.frame"`

UpperFaceMovements Inner Brow Raiser, Outer Brow Raiser, Brow Lower, Cheek Raiser, Nose Wrinkle.

Headmovements Turn Right, Turn Left, Up, Down.

MouthMovements Upper Lip Raiser, Nasolabial Furrow, Lip Puller, Lower Lip Depressor, Lip Pucker, Tongue Show, Lip Stretch, Lip Presser, Lip Suck, Lips Parting, Jaw Drop, Mouth Stretch.

TouchMovements Upper Face, Side Face, Lower Face, Mouth Area.

EyeBlink All scans were coded for eye blink.

status_bi "during the pandemic" is coded by 1, "before the pandemic" is coded by 0.

status specifies whether it is during or before the pandemic.

References

Reissland, N., Ustun, B. and Einbeck, J. (2024). The effects of lockdown during the COVID-19 pandemic on fetal movement profiles. BMC Pregnancy and Childbirth, 24(1), 1-7.

Examples

```
data(fetal_covid_data)
head(fetal_covid_data)
```

IALS_data   *International Adult Literacy Survey (IALS) for 13 countries*

Description

The data is obtained from the International Adult Literacy Survey (IALS), collected in 13 countries on Prose, Document, and Quantitative scales between 1994 and 1995. The data are reported as the percentage of individuals who could not reach a basic level of literacy in each country.

Usage

```
data(IALS_data)
```

Format

An object of class "data.frame"

Prose On prose scale, the percentage of individuals who could not reach a basic level of literacy in each country.

Document On document scale, the percentage of individuals who could not reach a basic level of literacy in each country.

Quantitative On quantitative scale, the percentage of individuals who could not reach a basic level of literacy in each country.

Country Specify the country

Gender Specify the gender

References

Sofroniou, N., Hoad, D., & Einbeck, J. (2008). League tables for literacy survey data based on random effect models. In: Proceedings of the 23rd International Workshop on Statistical Modelling, Utrecht; pp. 402-405.

Examples

```
data(IALS_data)
head(IALS_data)
```

---

mult.em_1level   *EM algorithm for multivariate one level model with covariates*

Description

This function is used to obtain the Maximum Likelihood Estimates (MLE) using the EM algorithm for one-level multivariate data. The estimates enable users to conduct clustering, ranking, and simultaneous dimension reduction on the multivariate dataset. Furthermore, when covariates are included, the function supports the fitting of multivariate response models, expanding its utility for regression analysis. The details of the model used in this function can be found in Zhang and Einbeck (2024). Note that this function is designed for multivariate data. When the dimension of the data is 1, please use alldist as an alternative. A warning message will also be displayed when the input data is a univariate dataset.

Arguments

data	A data set object; we denote the dimension to be m.
v	Covariate(s).
K	Number of mixture components, the default is K = 2. Note that when K = 1, z and beta will be 0.
steps	Number of iterations, the default is steps = 20.
start	Containing parameters involved in the proposed model (p, alpha, z, beta, sigma, gamma) in a list, the starting values can be obtained through the use of start_em. More details can be found in start_em.
option	Four options for selecting the starting values for the parameters in the model. The default is option = 1. More details can be found in start_em.
var_fun	There are four types of variance specifications; var_fun = 1, the same diagonal variance specification to all K components of the mixture; var_fun = 2, different diagonal variance matrices for different components. var_fun = 3, the same full (unrestricted) variance for all components. var_fun = 4, different full (unrestricted) variance matrices for different components. The default is var_fun = 2.

Value

The estimated parameters in the model $x_i = \alpha + \beta z_k + \Gamma v_i + \varepsilon_i$ obtained through the EM algorithm at the convergence.

p	The estimates for the parameter π_k, which is a vector of length K.
alpha	The estimates for the parameter α, which is a vector of length m.
z	The estimates for the parameter z_k, which is a vector of length K.
beta	The estimates for the parameter β, which is a vector of length m.
gamma	The estimates for the parameter Γ, which is a matrix.

sigma	The estimates for the parameter Σ_k. When var_fun = 1, Σ_k is a diagonal matrix and $\Sigma_k = \Sigma$, and we obtain a vector of the diagonal elements; When var_fun = 2, Σ_k is a diagonal matrix, and we obtain K vectors of the diagonal elements; When var_fun = 3, Σ_k is a full variance-covariance matrix, $\Sigma_k = \Sigma$, and we obtain a matrix Σ; When var_fun = 4, Σ_k is a full variance-covariance matrix, and we obtain K different matrices Σ_k.
W	The posterior probability matrix.
loglikelihood	The approximated log-likelihood of the fitted model.
disparity	The disparity (-2logL) of the fitted model.
number_parameters	
	The number of parameters estimated in the EM algorithm.
AIC	The AIC value (-2logL + 2number_parameters).
BIC	The BIC value (-2logL + number_parameters*log(n)), where n is the number of observations.
starting_values	
	A list of starting values for parameters used in the EM algorithm.

Note

It is worth noting that due to the sequential nature of the updates within the M-step, this algorithm can be considered an ECM algorithm.

References

Zhang, Y. and Einbeck, J. (2024). A Versatile Model for Clustered and Highly Correlated Multivariate Data. J Stat Theory Pract 18(5).doi:10.1007/s42519023003570

See Also

mult.reg_1level.

Examples

```
##example for data without covariates.
data(faithful)
res <- mult.em_1level(faithful,K=2,steps = 10,var_fun = 1)


## Graph showing the estimated one-dimensional space with cluster centers in red and alpha in green.
x <- res$alpha[1]+res$beta[1]*res$z
y <- res$alpha[2]+res$beta[2]*res$z
plot(faithful,col = 8)
points(x=x[1],y=y[1],type = "p",col = "red",pch = 17)
points(x=x[2],y=y[2],type = "p",col = "red",pch = 17)
points(x=res$alpha[1],y=res$alpha[2],type = "p",col = "darkgreen",pch = 4)
slope <- (y[2]-y[1])/(x[2]-x[1])
intercept <- y[1]-slope*x[1]
abline(intercept, slope, col="red")
```

```
##Graph showing the originaldata points being assigned to different
 ##clusters according to the Maximum a posterior (MAP) rule.
index <- apply(res$W, 1, which.max)
faithful_grouped <- cbind(faithful,index)
colors <- c("#FDAE61", "#66BD63")
plot(faithful_grouped[,-3], pch = 1, col = colors[factor(index)])


##example for data with covariates.
data(fetal_covid_data)
set.seed(2)
covid_res <- mult.em_1level(fetal_covid_data[,c(1:5)],v=fetal_covid_data$status_bi, K=3, steps = 20,
              var_fun = 2)
coeffs <- covid_res$gamma
##compare with regression coefficients from fitting individual linear models.
summary(lm( UpperFaceMovements ~ status_bi,data=fetal_covid_data))$coefficients[2,1]
summary(lm( Headmovements ~ status_bi,data=fetal_covid_data))$coefficients[2,1]
```

---

mult.em_2level   *EM algorithm for multivariate two level model with covariates*

Description

This function extends the one-level version mult.em_1level, and it is designed to obtain Maximum
Likelihood Estimates (MLE) using the EM algorithm for nested (structured) multivariate data, e.g.
multivariate test scores (such as on numeracy, literacy) of students nested in different classes or
schools. The resulting estimates can be applied for clustering or constructing league tables (ranking
of observations). With the inclusion of covariates, the model allows fitting a multivariate response
model for further regression analysis. Detailed information about the model used in this function
can be found in Zhang et al. (2023). Note that this function is designed for multivariate data. When
the dimension of the data is 1, please use allvc as an alternative. A warning message will also be
displayed when the input data is a univariate dataset.

Arguments

data	A data set object; we denote the dimension to be m.
v	Covariate(s).
K	Number of mixture components, the default is K = 2.
steps	Number of iterations, the default is steps = 20.
start	Containing parameters involved in the proposed model (p, alpha, z, beta, sigma, gamma) in a list, the starting values can be obtained through the use of start_em. More details can be found in start_em.
option	Four options for selecting the starting values for the parameters in the model. The default is option = 1. More details can be found in start_em.
var_fun	There are two types of variance specifications; var_fun = 1, the same diagonal variance specification to all K components of the mixture; var_fun = 2, different diagonal variance matrices for different components; The default is var_fun = 2.

Value

The estimated parameters in the model $x_{ij} = \alpha + \beta z_k + \Gamma v_{ij} + \varepsilon_{ij}$ obtained through the EM algorithm, where the upper-level unit is indexed by i, and the lower-level unit is indexed by j.

p	The estimates for the parameter π_k, which is a vector of length K.
alpha	The estimates for the parameter α, which is a vector of length m.
z	The estimates for the parameter z_k, which is a vector of length K.
beta	The estimates for the parameter β, which is a vector of length m.
gamma	The estimates for the parameter Γ, which is a matrix.
sigma	The estimates for the parameter Σ_k. When var_fun = 1, Σ_k is a diagonal matrix and $\Sigma_k = \Sigma$, and we obtain a vector of the diagonal elements; When var_fun = 2, Σ_k is a diagonal matrix, and we obtain K vectors of the diagonal elements.
W	The posterior probability matrix.
loglikelihood	The approximated log-likelihood of the fitted model.
disparity	The disparity (-2logL) of the fitted model.
number_parameters	
	The number of parameters estimated in the EM algorithm.
AIC	The AIC value (-2logL + 2number_parameters).
starting_values	
	A list of starting values for parameters used in the EM algorithm.

Note

It is worth noting that due to the sequential nature of the updates within the M-step, this algorithm can be considered an ECM algorithm.

References

Zhang, Y., Einbeck, J. and Drikvandi, R. (2023). A multilevel multivariate response model for data with latent structures. In: Proceedings of the 37th International Workshop on Statistical Modelling, pages 343-348. Link on RG: https://www.researchgate.net/publication/375641972_A_multilevel_multivariate_response_model_for_data_with_latent_structures

See Also

mult.reg_2level.

Examples

```
##examples for data without covariates.
data(trading_data)
set.seed(49)
trade_res <- mult.em_2level(trading_data, K=4, steps = 10, var_fun = 2)

i_1 <- apply(trade_res$W, 1, which.max)
ind_certain <- rep(as.vector(i_1),c(4,5,5,3,5,5,4,4,5,5,5,5,5,5,5,5,5,5,
3,5,5,5,5,4,4,5,5,5,4,5,4,5,5,5,3,5,5,5,5,5,5,4,5,4))
```

```
colors <- c("#FF6600","#66BD63", "lightpink","purple")
plot(trading_data[,-3],pch = 1, col = colors[factor(ind_certain)])
legend("topleft", legend=c("Mass point 1", "Mass point 2","Mass point 3","Mass point 4"),
col=c("#FF6600","purple","#66BD63","lightpink"),pch = 1, cex=0.8)

###The Twins data
library(lme4)
set.seed(26)
twins_res <- mult.em_2level(twins_data[,c(1,2,3)],v=twins_data[,c(4,5,6)],
K=2, steps = 20, var_fun = 2)
coeffs <- twins_res$gamma
##Compare to the estimated coefficients obtained using individual two-level models (lmer()).
summary(lmer(SelfTouchCodable ~ Depression + PSS + Anxiety + (1 | id) ,
data=twins_data, REML = TRUE))$coefficients[2,1]
```

---

mult.latent.reg   *Regression and Clustering in Multivariate Response Scenarios*

Description

This package implements methodology for the estimation of multivariate response models with random effects on one or two levels; whereby the (one-dimensional) random effect represents a latent variable approximating the multivariate space of outcomes, after possible adjustment for covariates. The estimation methodology makes use of a nonparametric maximum likelihood-type approach, where the random effect distribution is approximated by a discrete mixture, hence allowing the use of the EM algorithm for the estimation of all model parameters. The method is particularly useful for multivariate, highly correlated outcome variables with unobserved heterogeneities. Applications include regression with multivariate responses, as well as multivariate clustering or ranking problems. The details of the models can be found in Zhang and Einbeck (2024) and Zhang et al. (2023). The main functions are `mult.em_1level` and `mult.em_2level` for the fitting of the raw models, as well as envelope functions `mult.reg_1level` and `mult.reg_2level` which facilitate iterative runs of the algorithm with a view to finding optimal starting points, with help by function `start_em`.

Details

Package: mult.latent.reg

Type: Package

License: GPL-3

Author(s)

Yingjuan Zhang <yingjuan.zhang@durham.ac.uk>

Jochen Einbeck

References

Zhang, Y., Einbeck, J., and Drikvandi, R. (2023). A multilevel multivariate response model for data with latent structures. In: Proceedings of the 37th International Workshop on Statistical Modelling, Dortmund; pages 343-348. Link on RG: `https://www.researchgate.net/publication/375641972_A_multilevel_multivariate_response_model_for_data_with_latent_structures`.

Zhang, Y. and Einbeck, J. (2024). A Versatile Model for Clustered and Highly Correlated Multivariate Data. J Stat Theory Pract 18(5).doi:10.1007/s42519023003570

---

mult.reg_1level	*Selecting the best results for multivariate one level model*

Description

This wrapper function runs multiple times the function mult.em_1level for fitting Zhang and Einbeck's (2024) multivariate response models with one-level random effect, and select the best results with the smallest AIC value.

Arguments

data	A data set object; we denote the dimension of a data set to be m.
v	Covariate(s).
K	Number of mixture components, the default is K = 2.
steps	Number of iterations within each num_runs, the default is steps = 20.
num_runs	Number of function iteration runs, the default is num_runs = 10.
start	Containing parameters involved in the proposed model (p, alpha, z, beta, sigma, gamma) in a list, the starting values can be obtained through the use of start_em. More details can be found in start_em.
option	Four options for selecting the starting values for the parameters in the model. The default is option = 1. More details can be found in start_em.
var_fun	There are four types of variance specifications; var_fun = 1, the same diagonal variance specification to all K components of the mixture; var_fun = 2, different diagonal variance matrices for different components. var_fun = 3, the same full (unrestricted) variance for all components. var_fun = 4, different full (unrestricted) variance matrices for different components. The default is var_fun = 2.

Value

The best estimated result (with the smallest AIC value) in the model (Zhang and Einbeck, 2024) $x_i = \alpha + \beta z_k + \Gamma v_i + \varepsilon_i$ obtained through the EM algorithm.

p	The estimates for the parameter π_k, which is a vector of length K.
alpha	The estimates for the parameter α, which is a vector of length m.
z	The estimates for the parameter z_k, which is a vector of length K.

beta	The estimates for the parameter β, which is a vector of length m.
gamma	The estimates for the parameter Γ, which is a matrix.
sigma	The estimates for the parameter Σ_k. When var_fun = 1, Σ_k is a diagonal matrix and $\Sigma_k = \Sigma$, and we obtain a vector of the diagonal elements; When var_fun = 2, Σ_k is a diagonal matrix, and we obtain K vectors of the diagonal elements; When var_fun = 3, Σ_k is a full variance-covariance matrix, $\Sigma_k = \Sigma$, and we obtain a matrix Σ; When var_fun = 4, Σ_k is a full variance-covariance matrix, and we obtain K different matrices Σ_k.
W	The posterior probability matrix.
loglikelihood	The approximated log-likelihood of the fitted model.
disparity	The disparity (-2logL) of the fitted model.
number_parameters	
	The number of parameters estimated in the EM algorithm.
AIC	The AIC value (-2logL + 2number_parameters).
BIC	The BIC value (-2logL + number_parameters*log(n)), where n is the number of observations.
aic_data	All AIC values in each run.
Starting_values	
	Lists of starting values for parameters used in each num_runs. It allows reproduction of the best result (obtained from mult.reg_1level) in a single run using mult.em_1level by setting start equal to the list of starting values that were used to obtain the best result in mult.reg_1level.

References

Zhang, Y. and Einbeck J. (2024). A Versatile Model for Clustered and Highly Correlated Multivariate Data. J Stat Theory Pract 18(5).doi:10.1007/s42519023003570

See Also

mult.em_1level.

Examples

```
##run the mult.em_1level() multiple times and select the best results with the smallest AIC value
set.seed(7)
results <- mult.reg_1level(fetal_covid_data[,c(1:5)],v=fetal_covid_data$status_bi,
K=3, num_runs = 5,steps = 20, var_fun = 2, option = 1)
##Reproduce the best result: the best result is the 5th run in the above example.
rep_best_result <- mult.em_1level(fetal_covid_data[,c(1:5)],
v=fetal_covid_data$status_bi,
K=3, steps = 20, var_fun = 2, option = 1,
start = results$Starting_values[[5]])
```

---

mult.reg_2level	*Selecting the best results for multivariate two level model*

Description

This wrapper function runs multiple times the function [mult.em_2level](#) for fitting Zhang et al.'s (2023) multivariate response models with two-level random effect, and select the best results with the smallest AIC value.

Arguments

data	A data set object; we denote the dimension of a data set to be m.
v	Covariate(s).
K	Number of mixture components, the default is K = 2.
steps	Number of iterations within each num_runs, the default is steps = 20.
num_runs	Number of function iteration runs, the default is num_runs = 20.
start	Containing parameters involved in the proposed model (p, alpha, z, beta, sigma, gamma) in a list, the starting values can be obtained through the use of [start_em](#). More details can be found in [start_em](#).
option	Four options for selecting the starting values for the parameters in the model. The default is option = 1. More details can be found in [start_em](#).
var_fun	There are two types of variance specifications; var_fun = 1, the same diagonal variance specification to all K components of the mixture; var_fun = 2, different diagonal variance matrices for different components; The default is var_fun = 2.

Value

The best estimated result (with the smallest AIC value) in the model $x_{ij} = \alpha + \beta z_k + \Gamma v_{ij} + \varepsilon_{ij}$ obtained through the EM algorithm (Zhang et al., 2023), where the upper-level unit is indexed by i, and the lower-level unit is indexed by j.

p	The estimates for the parameter π_k, which is a vector of length K.
alpha	The estimates for the parameter α, which is a vector of length m.
z	The estimates for the parameter z_k, which is a vector of length K.
beta	The estimates for the parameter β, which is a vector of length m.
gamma	The estimates for the parameter Γ, which is a matrix.
sigma	The estimates for the parameter Σ_k. When var_fun = 1, Σ_k is a diagonal matrix and $\Sigma_k = \Sigma$, and we obtain a vector of the diagonal elements; When var_fun = 2, Σ_k is a diagonal matrix, and we obtain K vectors of the diagonal elements.
W	The posterior probability matrix.
loglikelihood	The approximated log-likelihood of the fitted model.
disparity	The disparity (-2logL) of the fitted model.

number_parameters

> The number of parameters estimated in the EM algorithm.

AIC   The AIC value (`-2logL + 2number_parameters`).

aic_data   All AIC values in each run.

Starting_values

> Lists of starting values for parameters used in each num_runs. It allows repro-
> duction of the best result (obtained from [mult.reg_2level](#)) in a single run using
> [mult.em_2level](#) by setting start equal to the list of starting values that were
> used to obtain the best result in [mult.reg_2level](#).

References

Zhang, Y., Einbeck, J. and Drikvandi, R. (2023). A multilevel multivariate response model for data
with latent structures. In: Proceedings of the 37th International Workshop on Statistical Modelling,
pages 343-348. Link on RG: https://www.researchgate.net/publication/375641972_A_multilevel_multivariate_response_model_for_data_with_latent_structures

See Also

[mult.em_2level](#).

Examples

```
##run the mult.em_2level() multiple times and select the best results with the smallest AIC value
set.seed(7)
results <- mult.reg_2level(trading_data, K=4, steps = 10, num_runs = 5,
                           var_fun = 2, option = 1)
## Reproduce the best result: the best result is the 2nd run in the above example.
rep_best_result <- mult.em_2level(trading_data, K=4, steps = 10,
var_fun = 2, option = 1,
start = results$Starting_values[[2]])
```

---

start_em   *Starting values for parameters*

Description

The starting values for parameters used for the EM algorithm in the functions: [mult.em_1level](#),
[mult.em_2level](#), [mult.reg_1level](#) and [mult.reg_2level](#).

Arguments

data   A data set object; we denote the dimension of a data set to be m.

v   Covariate(s); we denote the dimension of it to be r.

K   Number of mixture components, the default is K = 2.

steps	Number of iterations. This will only be used when using `option = 2` for both the 1-level model and the 2-level model. It should also be used when using `option = 3` and `option = 4` for the 1-level model, provided `var_fun` is set to either 3 or 4; the default is `steps = 20`.
option	Four options for selecting the starting values for the parameters. The default is `option = 1`. When `option = 1`: $\pi_k = \frac{1}{K}$, $z_k \sim$ rnorm(K, mean = 0, sd=1), α = column means, β = a random row minus alpha, Γ = coefficient estimates from separate linear models, Σ is diagonal matrix where the diagonals take the value of column standard deviations over K; when `option = 2`: use a short run (`steps = 5`) of the EM function which uses `option = 1` with `var_fun = 1` and use the estimates as the starting values for all the parameters; when `option = 3`: the starting value of β is the first principal component, and the starting values for the rest of the parameters are the same as described when `option = 1`; when `option = 4`: first, take the scores of the first principal component of the data and perform K-means, π_k is the proportion of the clustering assignments, and z_k take the values of the K-means centers, and the starting values for the rest of the parameters are the same as described when `option = 1`.
var_fun	The four variance specifications. When `var_fun = 1`, the same diagonal variance specification to all K components of the mixture; `var_fun = 2`, different diagonal variance matrices for different components. `var_fun = 3`, the same full (unrestricted) variance for all components. `var_fun = 4`, different full (unrestricted) variance matrices for different components. If unspecified, `var_fun = 2`. Note that for application propose, in two-level models, `var_fun` can only take values of 1 or 2.
p	optional; specifies starting values for π_k, it is input as a K-dimensional vector.
z	optional; specifies starting values for z_k, it is input as a K-dimensional vector.
beta	optional; specifies starting values for β, it is input as an m-dimensional vector.
alpha	optional; specifies starting values for α, it is input as an m-dimensional vector.
sigma	optional; specifies starting values for Σ_k (Σ, when `var_fun = 1` or `var_fun = 3`), when `var_fun = 1`, it is input as an m-dimensional vector, when `var_fun = 2`, it is input as a list (of length K) of m-dimensional vectors, when `var_fun = 3`, it is input as an $m \times m$ matrix, when `var_fun = 4`, it is input as a list (of length K) of $m \times m$ matrices.
gamma	optional; the coefficients for the covariates; specifies starting values for Γ, it is input as an $m \times r$ matrix.

Value

The starting values (in a list) for parameters in the models $x_i = \alpha + \beta z_k + \Gamma v_i + \varepsilon_i$ (Zhang and Einbeck, 2024) and $x_{ij} = \alpha + \beta z_k + \Gamma v_{ij} + \varepsilon_{ij}$ (Zhang et al., 2023) used in the four fucntions: mult.em_1level, mult.em_2level, mult.reg_1level and mult.reg_2level.

p	The starting value for the parameter π_k, which is a vector of length K.
alpha	The starting value for the parameter α, which is a vector of length m.
z	The starting value for the parameter z_k, which is a vector of length K.
beta	The starting value for the parameter β, which is a vector of length m.

gamma	The starting value for the parameter Γ, which is a matrix.
sigma	The starting value for the parameter Σ_k. When `var_fun = 1`, Σ_k is a diagonal matrix and $\Sigma_k = \Sigma$, and we obtain a vector of the diagonal elements; When `var_fun = 2`, Σ_k is a diagonal matrix, and we obtain K vectors of the diagonal elements; When `var_fun = 3`, Σ_k is a full variance-covariance matrix, $\Sigma_k = \Sigma$, and we obtain a matrix Σ; When `var_fun = 4`, Σ_k is a full variance-covariance matrix, and we obtain K different matrices Σ_k.

References

Zhang, Y., Einbeck, J. and Drikvandi, R. (2023). A multilevel multivariate response model for data with latent structures. In: Proceedings of the 37th International Workshop on Statistical Modelling, pages 343-348. Link on RG: `https://www.researchgate.net/publication/375641972_A_multilevel_multivariate_response_model_for_data_with_latent_structures`.

Zhang, Y. and Einbeck, J. (2024). A Versatile Model for Clustered and Highly Correlated Multivariate Data. J Stat Theory Pract 18(5).doi:10.1007/s42519023003570

Examples

```
##example for the faithful data.
data(faithful)
start <- start_em(faithful, option = 1)
```

---

trading_data	*A set of import and export data in 44 countries.*

Description

The variables are given as the percentage of imports and exports in relation to the overall GDP. The data set comprises data from 44 countries (for our analysis), we specifically selected the time period between 2018 and 2022.

Usage

```
data(trading_data)
```

Format

An object of class `"data.frame"`

import The fetus from the same twins share the same id number.

export frequency of self-touch for each fetus.

country frequency of twin-to-twin for each fetus.

Source

Trade in Goods and Services. `https://data.oecd.org/trade/trade-in-goods-and-services.htm`. Accessed on 2023-05-29.

Examples

```
data(trading_data)
head(trading_data)
```

---

twins_data	*A set of fetal movements data in twins.*

Description

This data was collected for research on the effects of maternal mental health on prenatal movements in twins and singletons (Reissland et al., 2021). There are two touch movement types of the fetus recorded: self-touch and twin-to-twin touch, and the mothers' mental health status was collected on three variables: depression, perceived stress scale and stress. There are 14 pairs of twins, 11 of the mothers were available for one scan and 3 of them were available for two scans, i.e. in total there are 34 observations. This dataset contains only the twins data from the original study.

Usage

```
data(twins_data)
```

Format

An object of class "data.frame"

id The fetus from the same twins share the same id number.

SelfTouchCodable frequency of self-touch for each fetus.

OtherTouchCodable frequency of twin-to-twin for each fetus.

Depression Depression scale of the mothers.

PSS Perceived Stress Scale of the mothers.

Anxiety Hospital Anxiety of the mothers.

References

Reissland, N., Einbeck, J., Wood, R., and Lane, A. (2021). Effects of maternal mental health on prenatal movement profiles in twins and singletons. Acta Paediatrica, 110(9):2553–2558.

Examples

```
data(twins_data)
head(twins_data)
```

Index