Package: macc (via r-universe)

September 12, 2024

Type Package

Title Mediation Analysis of Causality under Confounding

Version 1.0.1

Date 2017-08-20

Author Yi Zhao <yi_zhao@alumni.brown.edu>, Xi Luo
<xi.rossi.luo@gmail.com>

Maintainer Yi Zhao <yi_zhao@alumni.brown.edu>

Description Performs causal mediation analysis under confounding or correlated errors. This package includes a single level mediation model, a two-level mediation model, and a three-level mediation model for data with hierarchical structures. Under the two/three-level mediation model, the correlation parameter is identifiable and is estimated based on a hierarchical-likelihood, a marginal-likelihood or a two-stage method. See Zhao, Y., & Luo, X. (2014), Estimating Mediation Effects under Correlated Errors with an Application to fMRI, <arXiv:1410.7217> for details.

License GPL (>= 2)

Depends lme4, nlme, optimx, MASS, car

NeedsCompilation no

Repository CRAN

Date/Publication 2017-08-23 20:03:54 UTC

Contents

`macc-package` *Causal Mediation Analysis under Correlated Errors*

Description

macc performs causal mediation analysis under confounding or correlated errors. This package includes a single level mediation model, a two-level mediation model and a three-level mediation model for data with hierarchical structure. Under the two/three-level mediation model, the correlation parameter is identifiable and estimated based on a hierarchical-likelihood or a two-stage method.

Details

Package:	macc
Type:	Package
Version:	1.0.1
Date:	2017-08-20
License:	GPL (>=2)

Author(s)

Yi Zhao <yi_zhao@alumni.brown.edu> and Xi Luo <xi.rossi.luo@gmail.com>

Maintainer: Yi Zhao <yi_zhao@alumni.brown.edu>

References

Zhao, Y., & Luo, X. (2014). *Estimating Mediation Effects under Correlated Errors with an Application to fMRI.* arXiv preprint arXiv:1410.7217.

`env.single` *Simulated single-level dataset*

Description

"env.single" is an R environment containing a data frame of data generated from 100 trials, the true coefficients and the coavariance matrix of the model errors.

Usage

```
data("env.single")
```

Format

An R environment.

`data1` a data frame with `Z` the treatment assignment, `M` the mediator and `R` the interested outcome.

`Theta` a 2 by 2 matrix, which is the coefficients of the model.

`Sigma` a 2 by 2 matrix, which is the covariance matrix of the model errors.

Details

The number of subjects is 100. The coefficients are set to be $A = 0.5$, $C = 0.5$ and $B = -1$. The variances of the model errors are $\sigma_1^2 = 1$, $\sigma_2^2 = 4$ and the correlation is $\delta = 0.5$. See Section 5.1 of the reference.

References

Zhao, Y., & Luo, X. (2014). *Estimating Mediation Effects under Correlated Errors with an Application to fMRI.* arXiv preprint arXiv:1410.7217.

Examples

```
data(env.single)
dt<-get("data1",env.single)
```

`env.three` *Simulated three-level dataset*

Description

"env.three" is an R environment containing a data `list` generated from 50 subjects and 4 sessions, and the parameter settings used to generate the data.

Usage

```
data("env.three")
```

Format

An R environment.

`data3` a list of length 50, each contains a list of length 4 of a data frame with 3 variables.

`Theta` a 2 by 2 matrix, which is the value of the fixed effects.

`Sigma` the covariance matrix of the model error terms for the single level model.

`n` a 50 by 4 matrix, is the number of trials for each subject each session.

`Psi` the covariance matrix of the random effects in the mixed effects model.

`Lambda` the covariance matrix of the model errors in the mixed effects model.

`A` a 50 by 4 matrix, is the `A` value in the single-level for each subject each session.

`B` a 50 by 4 matrix, is the `B` value in the single-level for each subject each session.

`C` a 50 by 4 matrix, is the `C` value in the single-level for each subject each session.

Details

The number of subjects is $N = 50$ and the number of sessions is $K = 4$. Under each session of each subject, the number of trials is a random draw from a Poisson distribution with mean 100. The fixed effects are set to be $A = 0.5$, $C = 0.5$, and $B = -1$, and the variances of the model errors are $\sigma^2_{1_{ik}} = 1$, $\sigma^2_{2_{ik}} = 4$ and the correlation is $\delta = 0.5$. See Section 5.2 of the reference for details.

References

Zhao, Y., & Luo, X. (2014). *Estimating Mediation Effects under Correlated Errors with an Application to fMRI.* arXiv preprint arXiv:1410.7217.

Examples

```
data(env.three)
dt<-get("data3",env.three)
```

---

env.two *Simulated two-level dataset*

Description

"env.three" is an R environment containing a data list generated from 50 subjects, and the parameter settings used to generate the data.

Usage

```
data("env.two")
```

Format

An R environment.

`data2` a list of length 50, each contains a data frame with 3 variables.

`Theta` a 2 by 2 matrix, which is the population level model coefficients.

`Sigma` the covariance matrix of the model error terms for the single level model.

`n` a 50 by 1 matrix, is the number of trials for each subject.

`Lambda` the covariance matrix of the model errors in the coefficient regression model.

`A` a vector of length 50, is the `A` value in the single-level for each subject each session.

`B` a vector of length 50, is the `B` value in the single-level for each subject each session.

`C` a vector of length 50, is the `C` value in the single-level for each subject each session.

Details

The number of subjects is $N = 50$. For each subject, the number of trials is a random draw from a Poisson distribution with mean 100. The population level coefficients are set to be $A = 0.5$, $C = 0.5$ and $B = -1$, and the variances of the model errors are $\sigma^2_{1_i} = 1$, $\sigma^2_{2_i} = 4$ and the correlation is $\delta = 0.5$. See Section 5.2 of the reference for details. This is a special case of the three-level data with $K = 1$.

References

Zhao, Y., & Luo, X. (2014). *Estimating Mediation Effects under Correlated Errors with an Application to fMRI.* arXiv preprint arXiv:1410.7217.

Examples

```
data(env.two)
dt<-get("data2",env.two)
```

---

macc *Mediation Analysis of Causality under Confounding*

Description

This function performs causal mediation analysis under confounding or correlated errors for single level, two-level, and three-level mediation models.

Usage

```
macc(dat, model.type = c("single", "multilevel", "twolevel"),
  method = c("HL", "TS", "HL-TS"), delta = NULL, interval = c(-0.9, 0.9),
  tol = 0.001, max.itr = 500, conf.level = 0.95,
  optimizer = c("optimx", "bobyqa", "Nelder_Mead"), mix.pkg = c("nlme", "lme4"),
  random.indep = TRUE, random.var.equal = FALSE, u.int = FALSE, Sigma.update = TRUE,
  var.constraint = TRUE, random.var.update = TRUE, logLik.type = c("logLik", "HL"),
  error.indep = TRUE, error.var.equal = FALSE,
  sens.plot = FALSE, sens.interval = seq(-1, 1, by = 0.01), legend.pos = "topright",
  xlab = expression(delta), ylab = expression(hat(AB)), cex.lab = 1, cex.axis = 1,
  lgd.cex = 1, lgd.pt.cex = 1, plot.delta0 = TRUE, ...)
```

Arguments

`dat`	a data frame or a list of data. When it is a data frame, it contains `Z` as the treatment/exposure assignment, `M` as the mediator and `R` as the interested outcome and `model.type` should be "single". `Z`, `M` and `R` are all in one column. When it is a list, the list length is the number of subjects. For a two-level dataset, each list contains one data frame with `Z`, `M` and `R`, and `model.type` should be "twolevel"; for a three-level dataset, each subject list consists of `K` lists of data frame, and `model.type` should be "multilevel".
`model.type`	a character of model type, "single" for single level model, "multilevel" for three-level model and "twolevel" for two-level model.
`method`	a character of method that is used for the two/three-level mediation model. When `delta` is given, the method can be either "HL" (hierarchical-likelihood) or "TS" (two-stage); when `delta` is not given, the method can be "HL", "TS" or "HL-TS". The "HL-TS" method estimates `delta` by the "HL" method first and uses the "TS" method to estimate the rest parameters. For three-level model, when `method = "HL"` and `u.int = TRUE`, the parameters are estimated through a marginal-likelihood method.

`delta` a number gives the correlation between the model errors. Default value is `NULL`. When `model.type = "single"`, the default will be `0`. For two/three-level model, if `delta = NULL`, the value of `delta` will be estimated.

`interval` a vector of length two indicates the searching interval when estimating `delta`. Default is (-0.9,0.9).

`tol` a number indicates the tolerence of convergence for the "HL" method. Default is 0.001.

`max.itr` an integer indicates the maximum number of interation for the "HL" method. Default is 500.

`conf.level` a number indicates the significance level of the confidence intervals. Default is 0.95.

`optimizer` a character of the name of optimizing function(s) in the mixed effects model. This is used only for three-level model. For details, see `lmerControl`.

`mix.pkg` a character of the package used for the mixed effects model in a three-level mediation model.

`random.indep` a logic value indicates if the random effects in the mixed effects model are independent. Default is `TRUE` assuming the random effects are independent. This is used for `model.type = "multilevel"` only.

`random.var.equal` a logic value indicates if the variances of the random effects are identical. Default is `FALSE` assuming the variances are not identical. This is used for `model.type = "multilevel"` only.

`u.int` a logic value. Default is `FALSE`. When `u.int = TRUE`, a marginal-likelihood method, which integrates out the random effects in the mixed effects model, is used to estimate the parameters. This is used when `model.type = "multilevel"` and `method = "HL"` or `method = "HL-TS"`.

`Sigma.update` a logic value. Default is `TRUE`, and the estimated variances of the errors in the single level model will be updated in each iteration when running a two/three-level mediation model.

`var.constraint` a logic value. Default is `TRUE`, and an interval constraint is added on the variance components in the two/three-level mediation model.

`random.var.update` a logic value. Default is `TRUE`, and the estimates of the variance of the random effects in the mixed effects model are updated in each iteration. This is used when `model.type = "multilevel"` and `method = "HL"` or `method = "HL-TS"`.

`logLik.type` a character value indicating the type of likelihood value returned. It is used for "TS" method. When `logLik.type = "logLik"`, the log-likelihood of the mixed effects model is maximized; when `logLik.type = "HL"`, the summation of log-likelihood of the single level model and the mixed effects model is maximized. This is used for `model.type = "multilevel"`.

`error.indep` a logic value. Default is `TRUE`. This is used for `model.type = "twolevel"`. When `error.indep = TRUE`, the error terms in the three linear models for A, B and C are independent.

`error.var.equal`
: a logic value, Default is `FALSE`, This is used for `model.type = "twolevel"`. When `error.var.equal = TRUE`, the variances of the error terms in the three linear models for A, B and C are assumed to be identical.

`sens.plot`
: a logic value. Default is `FALSE`. This is used only for single level model. When `sens.plot = TRUE`, the sensitivity analysis will be performed and plotted.

`sens.interval`
: a sequence of `delta` values under which the sensitivity analysis is performed. Default is a sequence from -1 to 1 with increment 0.01. The elements with absolute value 1 will be excluded from the analysis.

`legend.pos`
: a character indicates the location of the legend when `sens.plot = TRUE`. This is used for single level model.

`xlab`
: a title for `x` axis in the sensitivity plot.

`ylab`
: a title for `y` axis in the sensitivity plot.

`cex.lab`
: the magnigication to be used for `x` and `y` labels relative to the current setting of `cex`.

`cex.axis`
: the magnification to be used for axis annotation relative to the current setting of `cex`.

`lgd.cex`
: the magnification to be used for legend relative to the current setting of `cex`.

`lgd.pt.cex`
: the magnification to be used for the points in legend relative to the current setting of `cex`.

`plot.delta0`
: a logic value. Default is `TRUE`. When `plot.delta0 = TRUE`, the estimates when $\delta = 0$ is plotted.

`...`
: additional argument to be passed.

Details

The single level mediation model is

$$M = ZA + E_1,$$

$$R = ZC + MB + E_2.$$

A correlation between the model errors E_1 and E_2 is assumed to be δ. The coefficients are estimated by maximizing the log-likelihood function. The confidence intervals of the coefficients are calculated based on the asymptotic joint distribution. The variance of AB estimator based on either the product method or the difference method is obtained from the Delta method. Under this single level model, δ is not identifiable. Sensitivity analysis for the indirect effect (AB) can be used to assess the deviation of the findings, when assuming $\delta = 0$ violates the independence assumption.

The two/three-level mediation models are proposed to estimate δ from data without sensitivity analysis. They address the within/between-subject variation issue for datasets with hierarchical structure. For simplicity, we refer to the three levels of data by trials, sessions and subjects, respectively. See reference for more details. Under the two-level mediation model, the data consists of N independent subjects and n_i trials for subject i; under the three-level mediation model, the data consists of N independent subjects, K sessions for each and n_{ik} trials. Under the two-level (three-level) models, the single level mediation model is first applied on the trials from (the same session of) a single subject. The coefficients then follow a linear (mixed effects) model. Here we enforce the assumption that δ is a constant across (sessions and) subjects. The parameters are estimated through a hierarchical-likelihood (or marginal likelihood) or a two-stage method.

Value

When `model.type = "single"`,

`Coefficients` point estimate of the coefficients, as well as the corresponding standard error and confidence interval. The indirect effect is estimated by both the product (`ABp`) and the difference (`ABd`) methods.

`D` point estimate of the regression coefficients in matrix form.

`Sigma` estimated covariance matrix of the model errors.

`delta` the δ value used to estimate the rest parameters.

`time` the CPU time used, see `system.time`.

When `model.type = "multilevel"`,

`delta` the specified or estimated value of correlation.

`Coefficients` the estimated fixed effects in the mixed effects model for the coefficients, as well as the corresponding confidence intervals and standard errors. Here confidence intervals and standard errors are the estimates directly from the mixed effects model, the variation of estimating these parameters is not accounted for.

`Cor.comp` estimated correlation matrix of the random effects in the mixed effects model.

`Var.comp` estimated variance components in the mixed effects model.

`Var.C2` estimated variance components for C', the total effect, if a mixed effects model is considered.

`logLik` the value of maximized log-likelihood of the mixed effects model.

`HL` the value of hierarchical-likelihood.

`convergence` the logic value indicating if the method converges.

`time` the CPU time used, see `system.time`

.

When `model.type = "twolevel"`

`delta` the specified or estimated value of correlation.

`Coefficients` the estimated population level effect in the regression models.

`Lambda` the estimated covariance matrix of the model errors in the coefficient regression models.

`Sigma` the estimated variances of E_1 and E_2 for each subject.

`HL` the value of full-likelihood (hierarchical-likelihood).

`convergence` the logic value indicating if the method converges.

`Var.constraint` the interval constraints used for the variances in the coefficient regression models.

`time` the CPU time used, see `system.time`

.

Author(s)

Yi Zhao, Brown University, <yi_zhao@brown.edu>; Xi Luo, Brown University, <xi.rossi.luo@gmail.com>.

References

Zhao, Y., & Luo, X. (2014). *Estimating Mediation Effects under Correlated Errors with an Application to fMRI.* arXiv preprint arXiv:1410.7217.

Examples

```
# Examples with simulated data

  ##############################################################################
  # Single level mediation model
 # Data was generated with 500 independent trials. The correlation between model errors is 0.5.
  data(env.single)
  data.SL<-get("data1",env.single)

  ## Example 1: Given delta is 0.5.
  macc(data.SL,model.type="single",delta=0.5)
  # $Coefficients
  #       Estimate        SE          LB          UB
  # A    0.3572722 0.1483680  0.06647618  0.64806816
  # C    0.8261253 0.2799667  0.27740060  1.37485006
  # B   -0.9260217 0.1599753 -1.23956743 -0.61247594
  # C2   0.4952836 0.2441369  0.01678400  0.97378311
  # ABp -0.3308418 0.1488060 -0.62249617 -0.03918738
  # ABd -0.3308418 0.3714623 -1.05889442  0.39721087

  ## Example 2: Assume the errors are independent.
  macc(data.SL,model.type="single",delta=0)
  # $Coefficients
  #          Estimate         SE          LB        UB
  # A    0.3572721688 0.14836803  0.06647618 0.6480682
  # C    0.4961424716 0.24413664  0.01764345 0.9746415
  # B   -0.0024040905 0.15997526 -0.31594984 0.3111417
  # C2   0.4952835570 0.24413691  0.01678400 0.9737831
  # ABp -0.0008589146 0.05715582 -0.11288227 0.1111644
  # ABd -0.0008589146 0.34526154 -0.67755910 0.6758413

  ## Example 3: Sensitivity analysis (given delta is 0.5).
  macc(data.SL,model.type="single",delta=0.5,sens.plot=TRUE)
  ##############################################################################

  ##############################################################################
  # Three-level mediation model
  # Data was generated with 50 subjects and 4 sessions.
  # The correlation between model errors in the single level is 0.5.
  # We comment out our examples due to the computation time.
  data(env.three)
  data.ML<-get("data3",env.three)
```

```
## Example 1: Correlation is unknown and to be estimated.
# Assume random effects are independent
# Add an interval constraint on the variance components.

# "HL" method
# macc(data.ML,model.type="multilevel",method="HL")
# $delta
# [1] 0.5224803

# $Coefficients
#            Estimate         LB         UB         SE
# A        0.51759400  0.3692202  0.6659678 0.07570229
# C        0.56882035  0.3806689  0.7569718 0.09599742
# B       -1.13624114 -1.3688690 -0.9036133 0.11868988
# C2      -0.06079748 -0.4163135  0.2947186 0.18138908
# AB.prod -0.58811160 -0.7952826 -0.3809406 0.10570145
# AB.diff -0.62961784 -1.0318524 -0.2273833 0.20522549
#
# $time
#    user  system elapsed
#   44.34    3.53   17.71

# "ML" method
# macc(data.ML,model.type="multilevel",method="HL",u.int=TRUE)
# $delta
# [1] 0.5430744

# $Coefficients
#            Estimate         LB         UB         SE
# A        0.51764821  0.3335094  0.7017871 0.09395011
# C        0.59652821  0.3715001  0.8215563 0.11481236
# B       -1.19426328 -1.4508665 -0.9376601 0.13092240
# C2      -0.06079748 -0.4163135  0.2947186 0.18138908
# AB.prod -0.61820825 -0.8751214 -0.3612951 0.13108056
# AB.diff -0.65732570 -1.0780742 -0.2365772 0.21467155
#
# $time
#    user  system elapsed
#  125.49    9.52   39.10

# "TS" method
# macc(data.ML,model.type="multilevel",method="TS")
# $delta
# [1] 0.5013719

# $Coefficients
#            Estimate         LB         UB         SE
# A        0.51805823  0.3316603  0.7044561 0.09510271
# C        0.53638546  0.3066109  0.7661601 0.11723409
# B       -1.07930526 -1.3386926 -0.8199179 0.13234293
# C2      -0.06079748 -0.4163135  0.2947186 0.18138908
# AB.prod -0.55914297 -0.8010745 -0.3172114 0.12343672
# AB.diff -0.59718295 -1.0204890 -0.1738769 0.21597645
```

```
#
# $time
#    user  system elapsed
#   19.53    0.00   19.54

## Example 2: Given the correlation is 0.5.
# Assume random effects are independent.
# Add an interval constraint on the variance components.

# "HL" method
macc(data.ML,model.type="multilevel",method="HL",delta=0.5)
# $delta
# [1] 0.5

# $Coefficients
#            Estimate         LB         UB         SE
# A        0.51760568  0.3692319  0.6659794 0.07570229
# C        0.53916412  0.3512951  0.7270331 0.09585330
# B       -1.07675116 -1.3093989 -0.8441035 0.11869999
# C2      -0.06079748 -0.4163135  0.2947186 0.18138908
# AB.prod -0.55733252 -0.7573943 -0.3572708 0.10207419
# AB.diff -0.59996161 -1.0020641 -0.1978591 0.20515811
#
# $time
#    user  system elapsed
#    2.44    0.22    1.03
################################################################################

################################################################################
# Two-level mediation model
# Data was generated with 50 subjects.
# The correlation between model errors in the single level is 0.5.
# We comment out our examples due to the computation time.
data(env.two)
data.TL<-get("data2",env.two)

## Example 1: Correlation is unknown and to be estimated.
# Assume errors in the coefficients regression models are independent.
# Add an interval constraint on the variance components.

# "HL" method
# macc(data.TL,model.type="twolevel",method="HL")
# $delta
# [1] 0.5066551

# $Coefficients
#            Estimate
# A        0.51714224
# C        0.54392056
# B       -1.05048406
# C2      -0.02924135
# AB.prod -0.54324968
# AB.diff -0.57316190
```

```
#
# $time
#    user  system elapsed
#    3.07    0.00    3.07

# "TS" method
# macc(data.TL,model.type="twolevel",method="TS")
# $delta
# [1] 0.4481611

# $Coefficients
#            Estimate
# A        0.52013697
# C        0.47945755
# B       -0.90252718
# C2      -0.02924135
# AB.prod -0.46943775
# AB.diff -0.50869890
#
# $time
#    user  system elapsed
#    1.60    0.00    1.59

## Example 2: Given the correlation is 0.5.
# Assume random effects are independent.
# Add an interval constraint on the variance components.

# "HL" method
macc(data.TL,model.type="twolevel",method="HL",delta=0.5)
# $delta
# [1] 0.5

# $Coefficients
#            Estimate
# A        0.51718063
# C        0.53543300
# B       -1.03336668
# C2      -0.02924135
# AB.prod -0.53443723
# AB.diff -0.56467434
#
# $time
#    user  system elapsed
#    0.21    0.00    0.20
##############################################################################
```

`sim.data.multi` *Generate two/three-level simulation data*

Description

This function generates a two/three-level dataset with given parameters.

Usage

```
sim.data.multi(Z.list, N, K = 1, Theta, Sigma,
    Psi = diag(rep(1, 3)), Lambda = diag(rep(1, 3)))
```

Arguments

`Z.list`	a list of data. When `K = 1` (a two-level dataset), each list is a vector containing the treatment/exposure assignment of the trials for each subject; When `K > 1` (a three-level dataset), each list is a list of length `K`, and each contains a vector of treatment/exposure assignment.
`N`	an integer, indicates the number of subjects.
`K`	an integer, indicates the number of sessions of each subject.
`Theta`	a 2 by 2 matrix, containing the population level model coefficients.
`Sigma`	a 2 by 2 matrix, is the covariance matrix of the model errors in the single-level model.
`Psi`	the covariance matrix of the random effects in the mixed effects model of the model coefficients. This is used only when `K > 1`. Default is a 3-dimensional identity matrix.
`Lambda`	the covariance matrix of the model errors in the mixed effects model if `K > 1` or the linear model if `K = 1` of the model coefficients.

Details

When `K > 1` (three-level data), for the n_{ik} trials in each session of each subject, the single level mediation model is

$$M_{ik} = Z_{ik}A_{ik} + E_{1ik},$$

$$R_{ik} = Z_{ik}C_{ik} + M_{ik}B_{ik} + E_{2ik},$$

where Z_{ik}, M_{ik}, R_{ik}, E_{1ik}, and E_{2ik} are vectors of length n_{ik}. `Sigma` is the covariance matrix of (E_{1ik}, E_{2ik}) (for simplicity, `Sigma` is the same across sessions and subjects). For coefficients A_{ik}, B_{ik} and C_{ik}, we assume a mixed effects model. The random effects are from a trivariate normal distribution with mean zero and covariance `Psi`; and the model errors are from a trivariate normal distribution with mean zero and covariance `Lambda`. For the fixed effects A, B and C, the values are specified in `Theta` with `Theta[1,1] = A`, `Theta[1,2] = C` and `Theta[2,2] = B`.

When `K = 1` (two-level data), the single-level model coefficients A_i, B_i and C_i are assumed to follow a trivariate linear regression model, where the population level coefficients are specified in `Theta` and the model errors are from a trivariate normal distribution with mean zero and covariance `Lambda`.

See Section 5.2 of the reference for details.

Value

`data`	a list of data. When `K = 1`, each list is a contains a dataframe; when `K > 1`, each list is a list of length `K`, and within each list is a dataframe.
`A`	the value of As. When `K = 1`, it is a vector of length `N`; when `K > 1`, it is a `N` by `K` matrix.

B	the value of Bs. When K = 1, it is a vector of length N; when K > 1, it is a N by K matrix.
C	the value of Cs. When K = 1, it is a vector of length N; when K > 1, it is a N by K matrix.
type	a character indicates the type of the dataset. When K = 1, type = twolevel; when K > 1, type = multilevel

Author(s)

Yi Zhao, Brown University, <yi_zhao@brown.edu>; Xi Luo, Brown University, <xi.rossi.luo@gmail.com>

References

Zhao, Y., & Luo, X. (2014). *Estimating Mediation Effects under Correlated Errors with an Application to fMRI.* arXiv preprint arXiv:1410.7217.

Examples

```
####################################################
 # Generate a two-level dataset

 # covariance matrix of errors
 delta<-0.5
 Sigma<-matrix(c(1,2*delta,2*delta,4),2,2)

 # model coefficients
 A0<-0.5
 B0<--1
 C0<-0.5

 Theta<-matrix(c(A0,0,C0,B0),2,2)

 # number of subjects, and trials of each
 set.seed(2000)
 N<-50
 K<-1
 n<-matrix(NA,N,K)
 for(i in 1:N)
 {
   n0<-rpois(1,100)
   n[i,]<-rpois(K,n0)
 }

 # treatment assignment list
 set.seed(100000)
 Z.list<-list()
 for(i in 1:N)
 {
   Z.list[[i]]<-rbinom(n[i,1],size=1,prob=0.5)
 }
```

```
# Lambda
rho.AB=rho.AC=rho.BC<-0
Lambda<-matrix(0,3,3)
lambda2.alpha=lambda2.beta=lambda2.gamma<-0.5
Lambda[1,2]=Lambda[2,1]<-rho.AB*sqrt(lambda2.alpha*lambda2.beta)
Lambda[1,3]=Lambda[3,1]<-rho.AC*sqrt(lambda2.alpha*lambda2.gamma)
Lambda[2,3]=Lambda[3,2]<-rho.BC*sqrt(lambda2.beta*lambda2.gamma)
diag(Lambda)<-c(lambda2.alpha,lambda2.beta,lambda2.gamma)

# Data
set.seed(5000)
re.dat<-sim.data.multi(Z.list=Z.list,N=N,K=K,Theta=Theta,Sigma=Sigma,Lambda=Lambda)
data2<-re.dat$data
###################################################

###################################################
# Generate a three-level dataset

# covariance matrix of errors
delta<-0.5
Sigma<-matrix(c(1,2*delta,2*delta,4),2,2)

# model coefficients
A0<-0.5
B0<--1
C0<-0.5

Theta<-matrix(c(A0,0,C0,B0),2,2)

# number of subjects, and trials of each
set.seed(2000)
N<-50
K<-4
n<-matrix(NA,N,K)
for(i in 1:N)
{
  n0<-rpois(1,100)
  n[i,]<-rpois(K,n0)
}

# treatment assignment list
set.seed(100000)
Z.list<-list()
for(i in 1:N)
{
  Z.list[[i]]<-list()
  for(j in 1:K)
  {
    Z.list[[i]][[j]]<-rbinom(n[i,j],size=1,prob=0.5)
  }
}

# Psi and Lambda
```

```
sigma2.alpha=sigma2.beta=sigma2.gamma<-0.5
theta2.alpha=theta2.beta=theta2.gamma<-0.5
Psi<-diag(c(sigma2.alpha,sigma2.beta,sigma2.gamma))
Lambda<-diag(c(theta2.alpha,theta2.beta,theta2.gamma))

# Data
set.seed(5000)
re.dat<-sim.data.multi(Z.list,N,K,Theta,Sigma,Psi,Lambda)
data3<-re.dat$data
###################################################
```

`sim.data.single` *Generate single-level simulation data*

Description

This function generates a single-level dataset with given parameters.

Usage

```
sim.data.single(Z, Theta, Sigma)
```

Arguments

`Z`	a vector of treatment/exposure assignment.
`Theta`	a 2 by 2 matrix, containing the model coefficients.
`Sigma`	a 2 by 2 matrix, is the covariance matrix of the model errors.

Details

The single level mediation model is

$$M = ZA + E_1,$$

$$R = ZC + MB + E_2.$$

`Theta[1,1]` = `A`, `Theta[1,2]` = `C` and `Theta[2,2]` = `B`; `Sigma` is the covariance matrix of (E_1, E_2).

Value

The function returns a dataframe with variables `Z`, `M` and `R`.

Author(s)

Yi Zhao, Brown University, <yi_zhao@brown.edu>; Xi Luo, Brown University, <xi.rossi.luo@gmail.com>

References

Zhao, Y., & Luo, X. (2014). *Estimating Mediation Effects under Correlated Errors with an Application to fMRI.* arXiv preprint arXiv:1410.7217.

Examples

```
##################################################
# Generate a single-level dataset

# covariance matrix of errors
delta<-0.5
Sigma<-matrix(c(1,2*delta,2*delta,4),2,2)

# model coefficients
A0<-0.5
B0<--1
C0<-0.5

Theta<-matrix(c(A0,0,C0,B0),2,2)

# number of trials
n<-100

# generate a treatment assignment vector
set.seed(100)
Z<-matrix(rbinom(n,size=1,0.5),n,1)

# Data
set.seed(5000)
data.single<-sim.data.single(Z,Theta,Sigma)
##################################################
```

Index