Package 'simPM'

March 3, 2019

Title SIMulation-based power analysis for Planned Missing designs

Version 0.0.0.9000

Description The package is being developed to automate the simulation-based power analysis for planned missing designs in the context of longitudinal studies. More specifically, this package is featured with a direct application to the scenarios where an unexpected funding cut occurs during the course of a study. Users can use this package to search for PM designs with sufficient power for testing focal parameters.

Depends MplusAutomation, lavaan, simsem, pheatmap, RColorBrewer

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

NeedsCompilation no

Author Yi Feng [aut, cre],
Gregory R. Hancock [aut]

Maintainer Yi Feng <yifeng94@umd.edu>

R topics documented:

`balance.miss` *Function to search for the indicator planned missing designs using Mplus*

Description

Function to search for the indicator planned missing designs using Mplus

Usage

```
balance.miss(VNAMES, distal.var = NULL, n, nreps, seed, Time, k,
  Time.complete, costmx, pc, pd, design0.out, focal.param,
  eval.budget = T, rm.budget = NULL, complete.var = NULL)
```

Arguments

`VNAMES`	The names of the observed variables, ordered by the time they are measured
`distal.var`	Any distal variables included in the model that would have complete data
`n`	The total sample size as initially planned
`nreps`	Number of replications for Monte Carlo simulation for each possible design
`seed`	Random seet for simulation
`Time`	The total number of time points (waves of data collection)
`k`	The number of observed variables collected at each wave
`Time.complete`	Number of waves of data collection that have been completed before the funding cut occured
`costmx`	The vector containing the unit cost of each observed variable which has no data collected yet. They are constant across subjects, but they can vary across variables and across time.
`pc`	The proportion of subjects that will participate in all of the following waves of data collection and provide complete data (must be greater than 0)
`pd`	The proportion of subjects that will not participate in any of the following waves of data collection (i.e., attritors). This can be 0.
`design0.out`	Mplus output file which contains the a priori power analysis/sample size planning (simulation) results for this specific model assuming a complete data design. Theoretically, such analysis was supposed to be conducted before the study began.
`focal.param`	The parameters of focal interest. If engine="l", the focal parameters should be specified using the lavaan script. If engine="m", the focal parameters should be specified based on the Mplus output file design0.out.
`eval.budget`	Budget constraints. If the researcher wishes to search for designs under the budget limit, you to provide the remaining available budget that can be used for data collection
`rm.budget`	The amount of remaining budget avaialbe for data collection
`complete.var`	Specify if there are any variables that need to have complete data collected across all the participating subjects

Value

An object containing the information of the optimal PM design, with highest power for testing the focal parameters, compared with other PM designs

See Also

simPM which wraps this function

`balance.miss.l` *Function to search for the optimal planned missing design using lavaan/simsem*

Description

Function to search for the optimal planned missing design using lavaan/simsem

Usage

```
balance.miss.l(popModel, analyzeModel, NAMES, Time, Time.complete, k, pc,
  pd, costmx, n, nreps, focal.param, complete.var = NULL,
  eval.budget = T, rm.budget = NULL, distal.var = NULL,
  seed = 1234)
```

Arguments

popModel	The data generation model (population model) specified using lavaan script
analyzeModel	The analysis model, which can be different from the population model, using lavaan script specific model assuming a complete data design. Theoretically, such analysis was supposed to be conducted before the study began.
NAMES	The names of the observed variables, ordered by the time they are measured
Time	The total number of time points (waves of data collection)
Time.complete	Number of waves of data collection that have been completed before the funding cut occured
k	The number of observed variables collected at each wave
pc	The proportion of subjects that will participate in all of the following waves of data collection and provide complete data (must be greater than 0)
pd	The proportion of subjects that will not participate in any of the following waves of data collection (i.e., attritors). This can be 0.
costmx	The vector containing the unit cost of each observed variable which has no data collected yet. They are constant across subjects, but they can vary across variables and across time.
n	The total sample size as initially planned
nreps	Number of replications for Monte Carlo simulation for each possible design
focal.param	The parameters of focal interest. If engine="l", the focal parameters should be specified using the lavaan script. If engine="m", the focal parameters should be specified based on the Mplus output file design0.out.
complete.var	Specify if there are any variables that need to have complete data collected across all the participating subjects

eval.budget	Budget constraints. If the researcher wishes to search for designs under the budget limit, you to provide the remaining available budget that can be used for data collection
rm.budget	The amount of remaining budget avaialbe for data collection
distal.var	Any distal variables included in the model that would have complete data
seed	Random seet for simulation

Value

An object containing the information of the optimal PM design, with highest power for testing the focal parameters, compared with other PM designs

See Also

simPM which wraps this function

`forward.opt` *Function to search for the optimal PM design using forward selection using Mplus*

Description

Function to search for the optimal PM design using forward selection using Mplus

Usage

```
forward.opt(VNAMES, distal.var, n, nreps, seed, Time, k, Time.complete,
  costmx, pc, pd, design0.out, focal.param, max.mk, eval.budget = F,
  rm.budget = NULL, complete.var = NULL)
```

Arguments

VNAMES	The names of the observed variables, ordered by the time they are measured
distal.var	Any distal variables included in the model that would have complete data
n	The total sample size as initially planned
nreps	Number of replications for Monte Carlo simulation for each possible design
seed	Random seet for simulation
Time	The total number of time points (waves of data collection)
k	The number of observed variables collected at each wave
Time.complete	Number of waves of data collection that have been completed before the funding cut occured
costmx	The vector containing the unit cost of each observed variable which has no data collected yet. They are constant across subjects, but they can vary across variables and across time.
pc	The proportion of subjects that will participate in all of the following waves of data collection and provide complete data (must be greater than 0)
pd	The proportion of subjects that will not participate in any of the following waves of data collection (i.e., attritors). This can be 0.

design0.out	Mplus output file which contains the a priori power analysis/sample size planning (simulation) results for this specific model assuming a complete data design. Theoretically, such analysis was supposed to be conducted before the study began.
focal.param	The parameters of focal interest. If engine="l", the focal parameters should be specified using the lavaan script. If engine="m", the focal parameters should be specified based on the Mplus output file design0.out.
max.mk	If using forward selection, specify the maximum number of unique missing data patterns in the design
eval.budget	Budget constraints. If the researcher wishes to search for designs under the budget limit, you to provide the remaining available budget that can be used for data collection
rm.budget	The amount of remaining budget avaialbe for data collection
complete.var	Specify if there are any variables that need to have complete data collected across all the participating subjects

Value

An object containing the information of the optimal PM design, with highest power for testing the focal parameters, compared with other PM designs

See Also

simPM which wraps this function

---

forward.opt.simsem *Function to search for the optimal PM design using lavaan/simsem*

Description

Function to search for the optimal PM design using lavaan/simsem

Usage

```
forward.opt.simsem(popModel, analyzeModel, NAMES, distal.var, n, nreps,
  seed, Time, k, Time.complete, costmx, pc, pd, focal.param, max.mk,
  eval.budget = F, rm.budget = NULL, complete.var = NULL)
```

Arguments

popModel	The data generation model (population model) specified using lavaan script
analyzeModel	The analysis model, which can be different from the population model, using lavaan script specific model assuming a complete data design. Theoretically, such analysis was supposed to be conducted before the study began.
NAMES	The names of the observed variables, ordered by the time they are measured
distal.var	Any distal variables included in the model that would have complete data
n	The total sample size as initially planned
nreps	Number of replications for Monte Carlo simulation for each possible design

seed	Random seet for simulation
Time	The total number of time points (waves of data collection)
k	The number of observed variables collected at each wave
Time.complete	Number of waves of data collection that have been completed before the funding cut occured
costmx	The vector containing the unit cost of each observed variable which has no data collected yet. They are constant across subjects, but they can vary across variables and across time.
pc	The proportion of subjects that will participate in all of the following waves of data collection and provide complete data (must be greater than 0)
pd	The proportion of subjects that will not participate in any of the following waves of data collection (i.e., attritors). This can be 0.
focal.param	The parameters of focal interest. If engine="l", the focal parameters should be specified using the lavaan script. If engine="m", the focal parameters should be specified based on the Mplus output file design0.out.
max.mk	If using forward selection, specify the maximum number of unique missing data patterns in the design
eval.budget	Budget constraints. If the researcher wishes to search for designs under the budget limit, you to provide the remaining available budget that can be used for data collection
rm.budget	The amount of remaining budget avaialbe for data collection
complete.var	Specify if there are any variables that need to have complete data collected across all the participating subjects

Value

An object containing the information of the optimal PM design, with highest power for testing the focal parameters, compared with other PM designs

See Also

simPM which wraps this function

---

opt.nm.1 *Select the optimal pattern with only one missing indicator using Mplus*

Description

Select the optimal pattern with only one missing indicator using Mplus

Usage

```
opt.nm.1(VNAMES, distal.var, n, nreps, seed, Time, k, Time.complete,
  costmx, pc, pd, design0.out, focal.param, complete.var = NULL)
```

Arguments

VNAMES	The names of the observed variables, ordered by the time they are measured
distal.var	Any distal variables included in the model that would have complete data
n	The total sample size as initially planned
nreps	Number of replications for Monte Carlo simulation for each possible design
seed	Random seet for simulation
Time	The total number of time points (waves of data collection)
k	The number of observed variables collected at each wave
Time.complete	Number of waves of data collection that have been completed before the funding cut occured
costmx	The vector containing the unit cost of each observed variable which has no data collected yet. They are constant across subjects, but they can vary across variables and across time.
pc	The proportion of subjects that will participate in all of the following waves of data collection and provide complete data (must be greater than 0)
pd	The proportion of subjects that will not participate in any of the following waves of data collection (i.e., attritors). This can be 0.
design0.out	Mplus output file which contains the a priori power analysis/sample size planning (simulation) results for this specific model assuming a complete data design. Theoretically, such analysis was supposed to be conducted before the study began.
focal.param	The parameters of focal interest. If engine="l", the focal parameters should be specified using the lavaan script. If engine="m", the focal parameters should be specified based on the Mplus output file design0.out.
complete.var	Specify if there are any variables that need to have complete data collected across all the participating subjects

Value

An object containing the information of the optimal PM design, with highest power for testing the focal parameters, compared with other PM designs

See Also

simPM which wraps this function

---

opt1.simsem *Function to search for the optimal missing pattern with one missing indicator*

Description

Function to search for the optimal missing pattern with one missing indicator

Usage

```
opt1.simsem(popModel, analyzeModel, NAMES, distal.var, n, nreps, seed,
  Time, k, Time.complete, costmx, pc, pd, focal.param,
  complete.var = NULL)
```

Arguments

popModel	The data generation model (population model) specified using lavaan script
analyzeModel	The analysis model, which can be different from the population model, using lavaan script specific model assuming a complete data design. Theoretically, such analysis was supposed to be conducted before the study began.
NAMES	The names of the observed variables, ordered by the time they are measured
distal.var	Any distal variables included in the model that would have complete data
n	The total sample size as initially planned
nreps	Number of replications for Monte Carlo simulation for each possible design
seed	Random seet for simulation
Time	The total number of time points (waves of data collection)
k	The number of observed variables collected at each wave
Time.complete	Number of waves of data collection that have been completed before the funding cut occured
costmx	The vector containing the unit cost of each observed variable which has no data collected yet. They are constant across subjects, but they can vary across variables and across time.
pc	The proportion of subjects that will participate in all of the following waves of data collection and provide complete data (must be greater than 0)
pd	The proportion of subjects that will not participate in any of the following waves of data collection (i.e., attritors). This can be 0.
focal.param	The parameters of focal interest. If engine="l", the focal parameters should be specified using the lavaan script. If engine="m", the focal parameters should be specified based on the Mplus output file design0.out.
complete.var	Specify if there are any variables that need to have complete data collected across all the participating subjects

Value

An object containing the information of the optimal PM design, with highest power for testing the focal parameters, compared with other PM designs

See Also

simPM which wraps this function

---

plotPM *Plot the missing data patterns*

Description

Plot the missing data patterns

Usage

```
plotPM(object, Time, k, colbr = "PRGn", col = c("antiquewhite1",
  "firebrick"))
```

Arguments

object	The simPM object
Time	The total number of waves
k	The number of observed variables at each wave
colbr	colors for waves
col	colors for complete vs. missing data

Examples

```
## Not run:
plotPM(wave.out,Time=5,k=3)
plotPM(indicator.out,Time=5,k=3)
plotPM(forward.out,Time=5,k=3)

## End(Not run)
```

---

pres.cost *To examine the lower level designs in forward selection*

Description

To examine the lower level designs in forward selection

Usage

```
pres.cost(opt.pattern, costmx, max.mk, pc, pd, n, k, Time, Time.complete)
```

Value

An object containing the information of the optimal PM design, with highest power for testing the focal parameters, compared with other PM designs

See Also

simPM which wraps this function

---

pres.design *To check the simulation results for lower level designs in forward selection*

Description

To check the simulation results for lower level designs in forward selection

Usage

```
pres.design(opt.results, max.mk)
```

Value

An object containing the information of the optimal PM design, with highest power for testing the focal parameters, compared with other PM designs

See Also

`simPM` which wraps this function

`simPM` *The wrapper function for all the searching strategies*

Description

The wrapper function for all the searching strategies

Usage

```
simPM(popModel, analyzeModel, design0.out = NULL, VarNAMES, Time,
  Time.complete, k, pc, pd, costmx, n, nreps, focal.param,
  complete.wave = NULL, complete.var = NULL, max.mk = NULL,
  eval.budget = T, rm.budget = NULL, distal.var = NULL,
  seed = 1234, engine = "l", methods = "wave")
```

Arguments

`popModel`	The data generation model (population model) specified using lavaan script
`analyzeModel`	The analysis model, which can be different from the population model, using lavaan script
`design0.out`	Mplus output file which contains the a priori power analysis/sample size planning (simulation) results for this specific model assuming a complete data design. Theoretically, such analysis was supposed to be conducted before the study began.
`VarNAMES`	The names of the observed variables, ordered by the time they are measured
`Time`	The total number of time points (waves of data collection)
`Time.complete`	Number of waves of data collection that have been completed before the funding cut occured
`k`	The number of observed variables collected at each wave
`pc`	The proportion of subjects that will participate in all of the following waves of data collection and provide complete data (must be greater than 0)
`pd`	The proportion of subjects that will not participate in any of the following waves of data collection (i.e., attritors). This can be 0.
`costmx`	The vector containing the unit cost of each observed variable which has no data collected yet. They are constant across subjects, but they can vary across variables and across time.
`n`	The total sample size as initially planned
`nreps`	Number of replications for Monte Carlo simulation for each possible design

focal.param	The parameters of focal interest. If engine="l", the focal parameters should be specified using the lavaan script. If engine="m", the focal parameters should be specified based on the Mplus output file design0.out.
complete.wave	Specify if there are any waves that need to have complete data collected across all the participating subjects
complete.var	Specify if there are any variables that need to have complete data collected across all the participating subjects
max.mk	If using forward selection, specify the maximum number of unique missing data patterns in the design
eval.budget	Budget constraints. If the researcher wishes to search for designs under the budget limit, you to provide the remaining available budget that can be used for data collection
rm.budget	The amount of remaining budget avaialbe for data collection
distal.var	Any distal variables included in the model that would have complete data
seed	Random seet for simulation
engine	Specify the whether the simulation should be completed using lavaan/simsem (engine="l") or Mplus (engine="m")
methods	Specify which searching method should be used ("wave","indicator","forward")

Value

An object containing the information of the optimal PM design, with highest power for testing the focal parameters, compared with other PM designs

See Also

simPM which wraps this function

---

summary.opt *A summary function to extract the important information of the output object.*

Description

A summary function to extract the important information of the output object.

Usage

```
summary.opt(object)
```

Value

An object containing the information of the optimal PM design, with highest power for testing the focal parameters, compared with other PM designs

See Also

simPM which wraps this function

wave.miss *Searching for wave missing designs, using Mplus*

Description

Searching for wave missing designs, using Mplus

Usage

```
wave.miss(VNAMES, distal.var = NULL, n, nreps, seed, Time, k,
  Time.complete, costmx, pc, pd, design0.out, focal.param,
  eval.budget = T, rm.budget = NULL, complete.wave = NULL)
```

Arguments

VNAMES	The names of the observed variables, ordered by the time they are measured
distal.var	Any distal variables included in the model that would have complete data
n	The total sample size as initially planned
nreps	Number of replications for Monte Carlo simulation for each possible design
seed	Random seet for simulation
Time	The total number of time points (waves of data collection)
k	The number of observed variables collected at each wave
Time.complete	Number of waves of data collection that have been completed before the funding cut occured
costmx	The vector containing the unit cost of each observed variable which has no data collected yet. They are constant across subjects, but they can vary across variables and across time.
pc	The proportion of subjects that will participate in all of the following waves of data collection and provide complete data (must be greater than 0)
pd	The proportion of subjects that will not participate in any of the following waves of data collection (i.e., attritors). This can be 0.
design0.out	Mplus output file which contains the a priori power analysis/sample size planning (simulation) results for this specific model assuming a complete data design. Theoretically, such analysis was supposed to be conducted before the study began.
focal.param	The parameters of focal interest. If engine="l", the focal parameters should be specified using the lavaan script. If engine="m", the focal parameters should be specified based on the Mplus output file design0.out.
eval.budget	Budget constraints. If the researcher wishes to search for designs under the budget limit, you to provide the remaining available budget that can be used for data collection
rm.budget	The amount of remaining budget avaialbe for data collection

Value

An object containing the information of the optimal PM design, with highest power for testing the focal parameters, compared with other PM designs

See Also

simPM which wraps this function

---

wave.miss.l *Searching for wave missing designs, using simsem/lavaan*

Description

Searching for wave missing designs, using simsem/lavaan

Usage

```
wave.miss.l(popModel, analyzeModel, NAMES, Time, Time.complete, k, pc, pd,
  costmx, n, nreps, focal.param, complete.wave = NULL, eval.budget = T,
  rm.budget = NULL, distal.var = NULL, seed = 1234)
```

Arguments

popModel	The data generation model (population model) specified using lavaan script
analyzeModel	The analysis model, which can be different from the population model, using lavaan script
NAMES	The names of the observed variables, ordered by the time they are measured
Time	The total number of time points (waves of data collection)
Time.complete	Number of waves of data collection that have been completed before the funding cut occured
k	The number of observed variables collected at each wave
pc	The proportion of subjects that will participate in all of the following waves of data collection and provide complete data (must be greater than 0)
pd	The proportion of subjects that will not participate in any of the following waves of data collection (i.e., attritors). This can be 0.
costmx	The vector containing the unit cost of each observed variable which has no data collected yet. They are constant across subjects, but they can vary across variables and across time.
n	The total sample size as initially planned
nreps	Number of replications for Monte Carlo simulation for each possible design
focal.param	The parameters of focal interest. If engine="l", the focal parameters should be specified using the lavaan script. If engine="m", the focal parameters should be specified based on the Mplus output file design0.out.
complete.wave	Specify if there are any waves that need to have complete data collected across all the participating subjects
eval.budget	Budget constraints. If the researcher wishes to search for designs under the budget limit, you to provide the remaining available budget that can be used for data collection
rm.budget	The amount of remaining budget avaialbe for data collection
distal.var	Any distal variables included in the model that would have complete data
seed	Random seet for simulation

Value

An object containing the information of the optimal PM design, with highest power for testing the focal parameters, compared with other PM designs

See Also

`simPM` which wraps this function

Index