Package 'simPM'

July 17, 2019

Title SIMulation-based power analysis for Planned Missing designs

Version 0.0.0.9000

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

Maintainer Yi Feng<yifeng94@umd.edu>

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.

URL https://yifengedms.github.io/simPM/

Depends MplusAutomation, lavaan, simsem, pheatmap, RColorBrewer

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

Suggests knitr,
rmarkdown

VignetteBuilder knitr

R topics documented:

`balance.miss` *Function to search for the item-level planned missing designs (via Mplus)*

Description

Function to search for the item-level planned missing designs (via 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`	A vector containing the names of the observed variables. The variable names must be ordered chronologically, by the time (wave) they are measured.d
`distal.var`	Specify the names of the variables, if there are any time-independent distal variables included in the model that are not subject planned missingness.
`n`	The total sample size as initially planned.
`nreps`	Number of replications for Monte Carlo simulations.
`seed`	seed for random number generation.
`Time`	The total number of time points (or 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 occurs.
`costmx`	A vector containing the unit cost of each observed variable that is yet to be measured (post the funding cut). The cost is assumed to be constant across subjects, but it is allowed to vary across variables and across waves.
`pc`	Proportion of completers. The proportion of subjects that will participate in all of the following waves of data collection and provide complete data. This 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., drop from the longitudinal study). This can be 0.
`design0.out`	Mplus output file which contains the a priori power analysis (i.e., simulations for sample size planning) 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. The focal parameters should be specified based on the Mplus output file design0.out.
`eval.budget`	Logical, indicating whether there is any budget constraint. If the user wishes to search for PM designs under the budget limit, they need to specify the amount of the remaining available budget that can be used for future data collection.
`rm.budget`	The amount of remaining budget avaialbe for future data collection.
`complete.var`	Specify the names of the variable(s) if there are any variable(s) 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 item-level PM design using lavaan/simsem*

Description

Function to search for the optimal item-level PM 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,specified using lavaan script. The analysis model can be different from the population model.
`NAMES`	A vector containing the names of the observed variables. The variable names must be ordered chronologically, by the time (wave) they are measured.
`Time`	The total number of time points (or waves of data collection).
`Time.complete`	Number of waves of data collection that have been completed before the funding cut occurs.
`k`	The number of observed variables collected at each wave.
`pc`	Proportion of completers. The proportion of subjects that will participate in all of the following waves of data collection and provide complete data. This 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., drop from the longitudinal study). This can be 0.
`costmx`	A vector containing the unit cost of each observed variable that is yet to be measured (post the funding cut). The cost is assumed to be constant across subjects, but it is allowed to vary across variables and across waves.
`n`	The total sample size as initially planned.
`nreps`	Number of replications for Monte Carlo simulations.
`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 the names of the variable(s) if there are any variable(s) that need to have complete data collected across all the participating subjects.
eval.budget	Logical, indicating whether there is any budget constraint. If the user wishes to search for PM designs under the budget limit, they need to specify the amount of the remaining available budget that can be used for future data collection.
rm.budget	The amount of remaining budget avaialbe for future data collection.
distal.var	Specify the names of the variables, if there are any time-independent distal variables included in the model that are not subject planned missingness.
seed	seed for random number generation.

Value

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

See Also

simPM which wraps this function

---

forward.opt *Function to search for the optimal PM design using forward selection (via Mplus)*

Description

Function to search for the optimal PM design using forward selection (via 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	A vector containing the names of the observed variables. The variable names must be ordered chronologically, by the time (wave) they are measured.
distal.var	Specify the names of the variables, if there are any time-independent distal variables included in the model that are not subject planned missingness.
n	The total sample size as initially planned.
nreps	Number of replications for Monte Carlo simulations.
seed	Random seet for simulation
Time	The total number of time points (or 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 occurs.

costmx	A vector containing the unit cost of each observed variable that is yet to be measured (post the funding cut). The cost is assumed to be constant across subjects, but it is allowed to vary across variables and across waves.
pc	Proportion of completers. The proportion of subjects that will participate in all of the following waves of data collection and provide complete data. This 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., drop from the longitudinal study). This can be 0.
design0.out	Mplus output file which contains the a priori power analysis (i.e., simulations for sample size planning) 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. The focal parameters should be specified #' based on the Mplus output file design0.out.
max.mk	Specify the maximum number of unique missing data patterns in the selected design. Only applicable if forward selection is used.
eval.budget	Logical, indicating whether there is any budget constraint. If the user wishes to search for PM designs under the budget limit, they need to specify the amount of the remaining available budget that can be used for future data collection.
rm.budget	The amount of remaining budget avaialbe for future data collection.
complete.var	Specify the names of the variable(s) if there are any variable(s) that need to have complete data collected across all the participating subjects.

Value

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

See Also

simPM which wraps this function

---

forward.opt.simsem *Function to search for the optimal PM design with forward selection (lavaan-based)*

Description

Function to search for the optimal PM design with forward selection (lavaan-based)

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,specified using lavaan script. The analysis model can be different from the population model.
NAMES	A vector containing the names of the observed variables. The variable names must be ordered chronologically, by the time (wave) they are measured.
distal.var	Specify the names of the variables, if there are any time-independent distal variables included in the model that are not subject planned missingness.
n	The total sample size as initially planned.
nreps	Number of replications for Monte Carlo simulations.
seed	Random seet for simulation
Time	The total number of time points (or 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 occurs.
costmx	A vector containing the unit cost of each observed variable that is yet to be measured (post the funding cut). The cost is assumed to be constant across subjects, but it is allowed to vary across variables and across waves.
pc	Proportion of completers. The proportion of subjects that will participate in all of the following waves of data collection and provide complete data. This 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., drop from the longitudinal study). This can be 0.
focal.param	The parameters of focal interest. The focal parameters should be specified #' using the lavaan script.
max.mk	Specify the maximum number of unique missing data patterns in the selected design. Only applicable if forward selection is used.
eval.budget	Logical, indicating whether there is any budget constraint. If the user wishes to search for PM designs under the budget limit, they need to specify the amount of the remaining available budget that can be used for future data collection.
rm.budget	The amount of remaining budget avaialbe for future data collection.
complete.var	Specify the names of the variable(s) if there are any variable(s) that need to have complete data collected across all the participating subjects.

Value

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

See Also

simPM which wraps this function

`opt.nm.1` *Function to search for the optimal missing pattern with one missing indicator. An internal function for forward selection.*

Description

Function to search for the optimal missing pattern with one missing indicator. An internal function for forward selection.

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`	A vector containing the names of the observed variables. The variable names must be ordered chronologically, by the time (wave) they are measured.
`distal.var`	Specify the names of the variables, if there are any time-independent distal variables included in the model that are not subject planned missingness.
`n`	The total sample size as initially planned.
`nreps`	Number of replications for Monte Carlo simulations.
`seed`	Random seet for simulation
`Time`	The total number of time points (or 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 occurs.
`costmx`	A vector containing the unit cost of each observed variable that is yet to be measured (post the funding cut). The cost is assumed to be constant across subjects, but it is allowed to vary across variables and across waves.
`pc`	Proportion of completers. The proportion of subjects that will participate in all of the following waves of data collection and provide complete data. This 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., drop from the longitudinal study). This can be 0.
`design0.out`	Mplus output file which contains the a priori power analysis (i.e., simulations for sample size planning) 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. The focal parameters should be specified based on the Mplus output file design0.out.
`complete.var`	Specify the names of the variable(s) if there are any variable(s) 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. An internal function for forward selection.*

Description

Function to search for the optimal missing pattern with one missing indicator. An internal function for forward selection.

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,specified using lavaan script. The analysis model can be different from the population model.
`NAMES`	A vector containing the names of the observed variables. The variable names must be ordered chronologically, by the time (wave) they are measured.
`distal.var`	Specify the names of the variables, if there are any time-independent distal variables included in the model that are not subject planned missingness.
`n`	The total sample size as initially planned.
`nreps`	Number of replications for Monte Carlo simulations.
`seed`	Random seet for simulation
`Time`	The total number of time points (or 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 occurs.
`costmx`	A vector containing the unit cost of each observed variable that is yet to be measured (post the funding cut). The cost is assumed to be constant across subjects, but it is allowed to vary across variables and across waves.
`pc`	Proportion of completers. The proportion of subjects that will participate in all of the following waves of data collection and provide complete data. This 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., drop from the longitudinal study). This can be 0.
`focal.param`	The parameters of focal interest. The focal parameters should be specified #' using the lavaan script.
`complete.var`	Specify the names of the variable(s) if there are any variable(s) 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 for the optimal PM design.*

Description

Plot the missing data patterns for the optimal PM design.

Usage

```
plotPM(object, colbr = "PRGn", col = c("antiquewhite1", "firebrick"),
  row.names = T, labels = T, fontsize_col = 20, fontsize_row = 14,
  fontsize = 14, angle_col = 45, legend = T, main = "")
```

Arguments

object	The simPM object
colbr	colors for waves. Default is "PRGn".
col	colors for complete vs. missing data. Default is c("antiquewhite1","firebrick").
labels	logical, indicating whether the label for waves is needed. Default is T.
fontsize_col	specify the font size for the column labels. Default is 20.
fontsize_row	specify the font size for the row labels. Default is 14.
fontsize	specify the font size for the legend. Default is 14.
angle_col	specify the angle of how the column labels are displayed
legend	logical, indicating whether the legend is shown. Default is T.
main	specify the plot title

See Also

pheatmap

Examples

```
## Not run:
plotPM(wave.out)
plotPM(indicator.out)
plotPM(forward.out,labels=F,col=c("gray96","gray35"),fontsize_row=26,fontsize=18,fontsize_col=26)

## 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,specified using lavaan script. The analysis model can be different from the population model.
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	A vector containing the names of the observed variables. The variable names must be ordered chronologically, by the time (wave) they are measured.
Time	The total number of time points (or waves of data collection).
Time.complete	Number of waves of data collection that have been completed before the funding cut occurs.
k	The number of observed variables collected at each wave.
pc	Proportion of completers. The proportion of subjects that will participate in all of the following waves of data collection and provide complete data. This 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., drop from the longitudinal study). This can be 0.
costmx	A vector containing the unit cost of each observed variable that is yet to be measured (post the funding cut). The cost is assumed to be constant across subjects, but it is allowed to vary across variables and across waves.
n	The total sample size as initially planned.
nreps	Number of replications for Monte Carlo simulations.
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 the wave(s) if there are any waves that need to have complete data collected across all the participants.
complete.var	Specify the names of the variable(s) if there are any variable(s) that need to have complete data collected across all the participating subjects.

max.mk	Specify the maximum number of unique missing data patterns in the selected design. Only applicable if forward selection is used.
eval.budget	Logical, indicating whether there is any budget constraint. If the user wishes to search for PM designs under the budget limit, they need to specify the amount of the remaining available budget that can be used for future data collection.
rm.budget	The amount of remaining budget avaialbe for future data collection.
distal.var	Specify the names of the variables, if there are any time-independent distal variables included in the model that are not subject planned missingness.
seed	seed for random number generation.
engine	Specify the whether the simulations should be conducted using lavaan/simsem (engine="l") or Mplus (engine="m").
methods	Specify which searching strategy 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

---

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

---

wave.miss *Searching for wave-level PM designs (Mplus based)*

Description

Searching for wave-level PM designs (Mplus based)

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	A vector containing the names of the observed variables. The variable names must be ordered chronologically, by the time (wave) they are measured.
distal.var	Specify the names of the variables, if there are any time-independent distal variables included in the model that are not subject planned missingness.
n	The total sample size as initially planned.
nreps	Number of replications for Monte Carlo simulations.
seed	seed for random number generation.
Time	The total number of time points (or 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 occurs.
costmx	A vector containing the unit cost of each observed variable that is yet to be measured (post the funding cut). The cost is assumed to be constant across subjects, but it is allowed to vary across variables and across waves.
pc	Proportion of completers. The proportion of subjects that will participate in all of the following waves of data collection and provide complete data. This 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., drop from the longitudinal study). This can be 0.
design0.out	Mplus output file which contains the a priori power analysis (i.e., simulations for sample size planning) 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. The focal parameters should be specified based on the Mplus output file design0.out.
eval.budget	Logical, indicating whether there is any budget constraint. If the user wishes to search for PM designs under the budget limit, they need to specify the amount of the remaining available budget that can be used for future data collection.
rm.budget	The amount of remaining budget avaialbe for future data collection.
complete.wave	Specify the wave(s) if there are any waves that need to have complete data collected across all the participants.

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 optimal wave-level PM designs (simsem/lavaan-based)*

Description

Searching for optimal wave-level PM designs (simsem/lavaan-based)

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,specified using lavaan script. The analysis model can be different from the population model.
`NAMES`	A vector containing the names of the observed variables. The variable names must be ordered chronologically, by the time (wave) they are measured.
`Time`	The total number of time points (or waves of data collection).
`Time.complete`	Number of waves of data collection that have been completed before the funding cut occurs.
`k`	The number of observed variables collected at each wave.
`pc`	Proportion of completers. The proportion of subjects that will participate in all of the following waves of data collection and provide complete data. This 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., drop from the longitudinal study). This can be 0.
`costmx`	A vector containing the unit cost of each observed variable that is yet to be measured (post the funding cut). The cost is assumed to be constant across subjects, but it is allowed to vary across variables and across waves.
`n`	The total sample size as initially planned.
`nreps`	Number of replications for Monte Carlo simulations.
`focal.param`	The parameters of focal interest. The focal parameters should be specified #' using the lavaan script.
`complete.wave`	Specify the wave(s) if there are any waves that need to have complete data collected across all the participants.
`eval.budget`	Logical, indicating whether there is any budget constraint. If the user wishes to search for PM designs under the budget limit, they need to specify the amount of the remaining available budget that can be used for future data collection.
`rm.budget`	The amount of remaining budget avaialbe for future data collection.
`distal.var`	Specify the names of the variables, if there are any time-independent distal variables included in the model that are not subject planned missingness.
`seed`	seed for random number generation.

Value

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

See Also

`simPM` which wraps this function

Index