

Package ‘simPM’

July 23, 2020

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 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 longitudinal study. Users can use this package to search for PM designs that yield sufficient power for testing focal parameters.

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

BugReports <https://github.com/YiFengEDMS/simPM/issues>

Depends R ($i= 3.5.0$),
lavaan ($i= 0.6-3$),
MplusAutomation ($i= 0.7-3$),
pheatmap ($i= 1.0.12$),
RColorBrewer ($i= 1.1-2$),
simsem ($i= 0.5-14$)

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

Suggests knitr,
rmarkdown,
testthat

VignetteBuilder knitr

R topics documented:

| | |
|------------------------------|---|
| balance.miss.1 | 2 |
| forward.opt.simsem | 3 |
| opt1.simsem | 5 |
| plotPM | 6 |
| pres.design | 7 |

| | |
|-------------------------|----|
| simPM | 7 |
| summary.simpm | 9 |
| wave.miss.l | 10 |

Index 12

| | |
|----------------|--|
| balance.miss.l | <i>Search for the optimal balanced item-level planned missing designs.</i> |
|----------------|--|

Description

balance.miss.l runs simulations using lavaan and simsem. It returns the search results for optimal balanced item-level PM designs.

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 character vector containing the names of the observed variables. The variable names must be ordered chronologically, by the time (wave) they are measured. |
| Time | Numeric. The total number of time points (or total number of waves of data collection). |
| Time.complete | Numeric. Number of waves of data collection that have been completed before the funding cut occurs. |
| k | Numeric. The number of observed variables collected at each wave. |
| pc | Numeric. 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 | Numeric. The proportion of subjects that will not participate in any of the following waves of data collection (i.e., people who will drop from the longitudinal study). This value can be 0. |
| costmx | A numeric 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 | Character vector. The parameters of focal interest. The focal parameters should be specified in the format of lavaan script. |
| complete.var | Char vector. Specify the name(s) of the variable(s) if there are any variable(s) that need to have complete data collected from all the participating subjects. |
| eval.budget | Logical scalar (TRUE or FALSE), 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 | Numeric. The amount of remaining budget available for future data collection. User must supply a value for this argument if eval.budget = TRUE. |
| distal.var | Char vector. Specify the name(s) of the distal variables. User needs to specify this argument if there are any time-independent distal variables included in the model that are not subject to planned missingness. |
| seed | The random seed for random number generation. |

Value

An object containing the information of the optimal balanced item-level missing design. The optimal design is the one that yields highest power for testing the focal parameters, compared to other plausible candidate PM designs.

See Also

[simPM](#) which is a wrapper function for this function.

forward.opt.simsem *Search for the optimal PM design using forward assembly.*

Description

forward.opt.simsem runs simulations using lavaan and simsem. It returns the search results for optimal item-level PM designs via forward assembly.

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 character vector containing the names of the observed variables. The variable names must be ordered chronologically, by the time (wave) they are measured. |

| | |
|----------------------------|---|
| <code>distal.var</code> | Char vector. Specify the name(s) of the distal variables. User needs to specify this argument if there are any time-independent distal variables included in the model that are not subject to planned missingness. |
| <code>n</code> | The total sample size as initially planned. |
| <code>nreps</code> | Number of replications for Monte Carlo simulations. |
| <code>seed</code> | The random seed for random number generation. |
| <code>Time</code> | Numeric. The total number of time points (or total number of waves of data collection). |
| <code>k</code> | Numeric. The number of observed variables collected at each wave. |
| <code>Time.complete</code> | Numeric. Number of waves of data collection that have been completed before the funding cut occurs. |
| <code>costmx</code> | A numeric 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. |
| <code>pc</code> | Numeric. 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. |
| <code>pd</code> | Numeric. The proportion of subjects that will not participate in any of the following waves of data collection (i.e., people who will drop from the longitudinal study). This value can be 0. |
| <code>focal.param</code> | Character vector. The parameters of focal interest. The focal parameters should be specified in the format of lavaan script. |
| <code>max.mk</code> | Specify the maximum number of unique missing data patterns in the selected design. Only applicable if forward assembly is used. |
| <code>eval.budget</code> | Logical scalar (TRUE or FALSE), 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. |
| <code>rm.budget</code> | Numeric. The amount of remaining budget available for future data collection. User must supply a value for this argument if <code>eval.budget = TRUE</code> . |
| <code>complete.var</code> | Char vector. Specify the name(s) of the variable(s) if there are any variable(s) that need to have complete data collected from all the participating subjects. |

Value

An object containing the information of the optimal item-level missing design. The optimal design is the one that yields highest power for testing the focal parameters, compared to other plausible candidate PM designs.

See Also

[simPM](#) which is a wrapper function for this function.

| | |
|-------------|---|
| opt1.simsem | <i>Search for the optimal missing pattern with only one missing measured variable. An internal function for forward assembly.</i> |
|-------------|---|

Description

opt1.simsem is an internal function that runs simulations using lavaan and simsem. It returns the optimal missing pattern that only contains one missing measured variable. This is the first step of forward assembly.

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 character vector containing the names of the observed variables. The variable names must be ordered chronologically, by the time (wave) they are measured. |
| distal.var | Char vector. Specify the name(s) of the distal variables. User needs to specify this argument if there are any time-independent distal variables included in the model that are not subject to planned missingness. |
| n | The total sample size as initially planned. |
| nreps | Number of replications for Monte Carlo simulations. |
| seed | The random seed for random number generation. |
| Time | Numeric. The total number of time points (or total number of waves of data collection). |
| k | Numeric. The number of observed variables collected at each wave. |
| Time.complete | Numeric. Number of waves of data collection that have been completed before the funding cut occurs. |
| costmx | A numeric 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 | Numeric. 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 | Numeric. The proportion of subjects that will not participate in any of the following waves of data collection (i.e., people who will drop from the longitudinal study). This value can be 0. |
| focal.param | Character vector. The parameters of focal interest. The focal parameters should be specified in the format of lavaan script. |

`complete.var` Char vector. Specify the name(s) of the variable(s) if there are any variable(s) that need to have complete data collected from all the participating subjects.

Value

An object containing the information of the optimal missing data pattern containing only one missing observed variable. The optimal pattern is the one that yields highest statistical power for testing the focal parameters, compared to other patterns with only one missing observed variable.

See Also

[simPM](#) which is a wrapper function for this function.

| | |
|---------------------|--|
| <code>plotPM</code> | <i>Plot the missing data patterns for the optimal PM design.</i> |
|---------------------|--|

Description

`plotPM` plots the optimal PM design's missing patterns as a heatmap.

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

| | |
|---------------------------|---|
| <code>object</code> | The <code>simPM</code> object. |
| <code>colbr</code> | Specify the colors for different waves. Default is "PRGn". |
| <code>col</code> | Specify the colors for complete vs. missing data. Default is <code>c("antiquewhite1", "firebrick")</code> . |
| <code>labels</code> | logical scalar, indicating whether the label for waves is needed. Default is TRUE. |
| <code>fontsize_col</code> | Specify the font size for the column labels. Default is 20. |
| <code>fontsize_row</code> | Specify the font size for the row labels. Default is 14. |
| <code>fontsize</code> | Specify the font size for the legend. Default is 14. |
| <code>angle_col</code> | Specify the angle of how the column labels are displayed. Default is 45. |
| <code>legend</code> | Logical scalar, indicating whether the legend is shown. Default is TRUE. |
| <code>main</code> | Specify the plot title. |
| <code>...</code> | Any additional arguments for pheatmap . |

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.design | <i>Check the simulation results for lower level designs in forward assembly. pres.design is only applicable when Mplus is used for simulations.</i> |
|-------------|---|

Description

Check the simulation results for lower level designs in forward assembly. pres.design is only applicable when Mplus is used for simulations.

Usage

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

Arguments

| | |
|-------------|--|
| opt.results | The object returned from forward assembly. |
| max.mk | The maximum number of unique missing data patterns in the selected design. |

| | |
|-------|---|
| simPM | <i>Search for an optimal PM design.</i> |
|-------|---|

Description

simPM runs Monte Carlo simulations and returns the search results for optimal PM designs. This is a wrapper function for all the available searching methods.

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 = "1", 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 | An object returned by <code>readModels</code> . To obtain this object, the user need to have a Mplus output file which contains the <i>a priori</i> power analysis results for this specific model assuming a complete data design (i.e., simulation-based power analysis for sample size planning). In principle, <i>a priori</i> power analysis is supposed to be conducted before the study began. |
| VarNAMES | A character vector containing the names of the observed variables. The variable names must be ordered chronologically, by the time (wave) they are measured. |
| Time | Numeric. The total number of time points (or total number of waves of data collection). |
| Time.complete | Numeric. Number of waves of data collection that have been completed before the funding cut occurs. |
| k | Numeric. The number of observed variables collected at each wave. |
| pc | Numeric. 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 | Numeric. The proportion of subjects that will not participate in any of the following waves of data collection (i.e., people who will drop from the longitudinal study). This value can be 0. |
| costmx | A numeric 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 | Character Vector. Specify the parameters of focal interest. If engine="l", the focal parameters should be specified following the format of lavaan script. If engine="m", the focal parameters should be specified in the specific format based on the Mplus output object <code>design0.out</code> . |
| complete.wave | Numeric vector. Specify which wave(s) that the user wish to have complete data collected from all the participants. Only applicable for wave-level PM designs. |
| complete.var | Char vector. Specify the name(s) of the variable(s) if there are any variable(s) that need to have complete data collected from all the participating subjects. |
| max.mk | Specify the maximum number of unique missing data patterns in the selected design. Only applicable if forward assembly is used. |
| eval.budget | Logical scalar (TRUE or FALSE), 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 | Numeric. The amount of remaining budget available for future data collection. User must supply a value for this argument if <code>eval.budget = TRUE</code> . |
| distal.var | Char vector. Specify the name(s) of the distal variables. User needs to specify this argument if there are any time-independent distal variables included in the model that are not subject to planned missingness. |
| seed | The random seed for random number generation. |
| engine | Specify the whether the simulations should be conducted using <code>lavaan/simsem</code> (<code>engine="l"</code>) or <code>Mplus</code> (<code>engine="m"</code>). |
| methods | Specify which searching strategy should be used (wave-level PM designs only: <code>methods = "wave"</code> ; balanced item-level PM designs only: <code>methods = "indicator"</code> ; item-level PM designs via forward assembly: <code>methods = "forward"</code>). |

Value

An object containing the information of the optimal PM design. The optimal design is the one that yields highest statistical power for testing the focal parameters, compared to other plausible candidate PM designs.

| | |
|---------------|---|
| summary.simpm | <i>Extract the summary information of the simpm object.</i> |
|---------------|---|

Description

`summary.simpm` summarizes and extracts the important information about the optimal PM design.

Usage

```
## S3 method for class 'simpm'
summary(object)
```

Arguments

`object` The output object returned by `simPM`.

Value

The information about the optimal PM design.

wave.miss.l *Search for the optimal wave-level PM designs.*

Description

wave.miss.l runs simulations using lavaan and simsem. It returns the search results for optimal wave-level PM designs.

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 character vector containing the names of the observed variables. The variable names must be ordered chronologically, by the time (wave) they are measured. |
| Time | Numeric. The total number of time points (or total number of waves of data collection). |
| Time.complete | Numeric. Number of waves of data collection that have been completed before the funding cut occurs. |
| k | Numeric. The number of observed variables collected at each wave. |
| pc | Numeric. 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 | Numeric. The proportion of subjects that will not participate in any of the following waves of data collection (i.e., people who will drop from the longitudinal study). This value can be 0. |
| costmx | A numeric 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 | Character vector. The parameters of focal interest. The focal parameters should be specified in the format of lavaan script. |
| complete.wave | Numeric vector. Specify which wave(s) that the user wish to have complete data collected from all the participants. |
| eval.budget | Logical scalar (TRUE or FALSE), 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. |

| | |
|-------------------------|---|
| <code>rm.budget</code> | Numeric. The amount of remaining budget available for future data collection. User must supply a value for this argument if <code>eval.budget = TRUE</code> . |
| <code>distal.var</code> | Char vector. Specify the name(s) of the distal variables. User needs to specify this argument if there are any time-independent distal variables included in the model that are not subject to planned missingness. |
| <code>seed</code> | The random seed for random number generation. |

Value

An object containing the information of the optimal wave-level missing design. The optimal design is the one that yields highest power for testing the focal parameters, compared to other plausible candidate PM designs.

See Also

[simPM](#) which is a wrapper function for this function.

Index

balance.miss.l, [2](#)
forward.opt.simsem, [3](#)
opt1.simsem, [5](#)
pheatmap, [6](#)
plotPM, [6](#)
pres.design, [7](#)
readModels, [8](#)
simPM, [3](#), [4](#), [6](#), [7](#), [9](#), [11](#)
summary.simpM, [9](#)
wave.miss.l, [10](#)