

Package ‘simPsyStudy’

September 15, 2026

Title Simulation of Ordinal Responses for Psychometric Studies

Version 1.1.8

Date 2026-09-07

Description Provides tools to define factorial simulation conditions and generate binary or ordinal item responses under common-factor and probit graded response model parameterizations. Supports multivariate normal and correlated gamma latent traits, reproducible replications, parameter conversion, and structured storage of generated datasets. The graded response model follows Samejima (1969).

License MIT + file LICENSE

RoxygenNote 7.2.3

Encoding UTF-8

Imports MASS, moments, MultiRNG, Rdpack

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

RdMacros Rdpack

URL <https://github.com/Boklauth/simPsyStudy>

BugReports <https://github.com/Boklauth/simPsyStudy/issues>

NeedsCompilation no

Author Bo Klauth [aut, cre] (ORCID: <<https://orcid.org/0000-0001-9442-4223>>)

Maintainer Bo Klauth <bklauth@gmail.com>

Repository CRAN

Date/Publication 2026-09-15 13:00:02 UTC

Contents

create_theta_mvn	2
diag_residuals	3
gamma_from_normal	3

gamma_from_normal2	5
simdata_faov	6
simdata_grm	7
simdata_grm2	8
simulate_condition	9
to_cormatrix	11
to_covmatrix	12
to_loadingmatrix	12
TSK	13

Index	14
--------------	-----------

create_theta_mvn	<i>Generate Multivariate Normal Latent Traits</i>
------------------	---

Description

Generates latent-trait values from a multivariate normal distribution and returns both the requested-scale and standardized values.

Usage

```
create_theta_mvn(  
  size,  
  mean_vec,  
  sd_vec,  
  ifcor_vec = numeric(0),  
  seed_num = 12345L  
)
```

Arguments

size	Positive integer sample size.
mean_vec	Finite numeric vector of marginal means.
sd_vec	Positive finite numeric vector of marginal standard deviations.
ifcor_vec	Pairwise correlations ordered as (1, 2), (1, 3), ..., (m - 1, m). Use numeric(0) for one dimension.
seed_num	Positive integer random-number seed.

Value

A list containing the generated matrix X, standardized matrix scaled.X, marginal skewness and kurtosis, and sample correlation and covariance matrices.

See Also

[to_cormatrix\(\)](#), [to_covmatrix\(\)](#)

Examples

```
theta <- create_theta_mvn(
  size = 100, mean_vec = c(0, 0), sd_vec = c(1, 1),
  ifcor_vec = 0.4, seed_num = 2026
)
head(theta$scaled.X)
```

diag_residuals

*Compute Residual Variances from Standardized Loadings***Description**

Compute Residual Variances from Standardized Loadings

Usage

```
diag_residuals(x)
```

Arguments

x A loading vector or item-by-factor loading matrix.

Value

A diagonal matrix containing 1 - communality for every item.

Examples

```
diag_residuals(c(0.7, 0.8, 0.6))
diag_residuals(to_loadingmatrix(rep(0.7, 4), list(1:2, 3:4)))
```

gamma_from_normal

*Generate a Skewed Multivariate Distribution by Rejection Sampling***Description**

Uses the package's original rejection-sampling approach to accept three-dimensional multivariate-normal draws according to gamma-to-normal density ratios.

Usage

```
gamma_from_normal(shape, rate, mean_vec, cov_matrix, size, c, seed_num)
```

Arguments

shape	Shape parameter for the gamma density.
rate	Rate parameter for the gamma density.
mean_vec	A numeric mean vector of length three for the multivariate normal distribution.
cov_matrix	A three-by-three covariance matrix for the multivariate normal distribution.
size	Positive integer sample size.
c	Positive rejection-sampling constant used in the gamma-to-normal density ratios.
seed_num	Positive integer random-number seed.

Details

This is the original simPsyStudy algorithm. Each iteration draws one three-dimensional multivariate-normal vector and one uniform value. The entire vector is retained when all three marginal rejection criteria are satisfied.

Value

A list containing the accepted values X , standardized values scaled.X , marginal skewness and kurtosis, and sample correlation and covariance matrices.

References

- Braun WJ, Murdoch DJ (2016). *A First Course in Statistical Programming with R*, 2nd edition. Cambridge University Press, Cambridge. ISBN 978-1-107-57646-9, doi:[10.1017/CBO9781316451090](https://doi.org/10.1017/CBO9781316451090).
- Rizzo ML (2007). *Statistical Computing with R*. Chapman and Hall/CRC, Boca Raton, FL. ISBN 978-1-58488-545-0.
- Robert CP, Casella G (2010). *Introducing Monte Carlo Methods with R*, 1st edition. Springer, New York, NY. ISBN 978-1-4419-1575-7, doi:[10.1007/9781441915764](https://doi.org/10.1007/9781441915764).

See Also

[stats::dgamma\(\)](#), [stats::dnorm\(\)](#), [gamma_from_normal2\(\)](#)

Examples

```
correlation <- to_cormatrix(c(0.49, 0.74, 0.87), n_dim = 3)
covariance <- to_covmatrix(correlation, c(0.95, 0.98, 1.1))
theta <- gamma_from_normal(
  shape = 5, rate = 5, mean_vec = c(-0.1, 0, 0.1),
  cov_matrix = covariance, size = 30, c = 4, seed_num = 45679
)
head(theta$scaled.X)
```

gamma_from_normal2	<i>Generate a Skewed Multivariate Gamma Distribution with a Gaussian Copula</i>
--------------------	---

Description

Generates correlated gamma variables by transforming multivariate-normal values through a Gaussian copula. This provides a second approach for generating positively skewed multivariate data.

Usage

```
gamma_from_normal2(shape, rate, mean_vec, cov_matrix, size, seed_num = 12345L)
```

Arguments

shape	Positive gamma shape value, or one value per dimension.
rate	Positive gamma rate value, or one value per dimension.
mean_vec	Finite mean vector for the latent multivariate-normal distribution.
cov_matrix	Positive-definite covariance matrix for the latent multivariate-normal distribution. Its correlation matrix controls the dependence of the generated gamma variables.
size	Integer sample size of at least two.
seed_num	Positive integer random-number seed.

Details

The gamma marginal distributions follow the requested shape and rate. The input covariance controls latent Gaussian dependence, so the output Pearson correlations generally differ from it.

Value

A list containing gamma values X , standardized values $\text{scaled}.X$, marginal skewness and kurtosis, and sample correlation and covariance matrices.

See Also

[gamma_from_normal\(\)](#)

Examples

```
latent_cor <- to_cormatrix(c(0.4, 0.3, 0.5), 3)
theta <- gamma_from_normal2(
  shape = 5, rate = 5, mean_vec = c(0, 0, 0),
  cov_matrix = latent_cor, size = 100, seed_num = 2026
)
head(theta$scaled.X)
```

simdata_faov

*Simulate Ordinal Responses from a Common-Factor Model***Description**

Simulates continuous response variables from a common-factor model and categorizes them using supplied thresholds. Latent scores can be supplied or generated from a multivariate normal distribution in every replication.

Usage

```
simdata_faov(
  model,
  f_mean,
  f_cov_matrix,
  theta_matrix = NULL,
  vloadings,
  thresholds,
  N = NULL,
  R,
  seed = 123456L,
  file_dir = tempdir(),
  file_prefix = "sim"
)
```

Arguments

<code>model</code>	A nonempty list assigning each item to exactly one factor.
<code>f_mean</code>	Finite vector containing one latent mean per factor.
<code>f_cov_matrix</code>	Positive-definite factor covariance matrix.
<code>theta_matrix</code>	Optional finite numeric matrix with people in rows and factors in columns. A vector is accepted for one factor. When supplied, these values remain fixed across replications.
<code>vloadings</code>	Finite vector containing one standardized loading per item.
<code>thresholds</code>	Numeric item-by-threshold matrix with values increasing across rows. A vector is accepted when there is one threshold per item.
<code>N</code>	Sample size. May be omitted when <code>theta_matrix</code> is supplied.
<code>R</code>	Positive integer number of replications.
<code>seed</code>	Positive integer base seed. Replication <code>r</code> uses <code>seed + r</code> .
<code>file_dir</code>	Directory in which to write generated datasets.
<code>file_prefix</code>	Filename prefix containing letters, digits, underscores, or hyphens.

Value

A list containing population parameters, response proportions, and the simulation specification.

References

Muthén B (1978). “Contributions to factor analysis of dichotomous variables.” *Psychometrika*, **43**(4), 551–560. ISSN 0033-3123, doi:10.1007/BF02293813.

Muthén B (1984). “A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators.” *Psychometrika*, **49**(1), 115–132. ISSN 0033-3123, doi:10.1007/BF02294210.

Examples

```
out_dir <- tempfile("simPsyStudy-faov-")
result <- simdata_faov(
  model = list(1:3), f_mean = 0, f_cov_matrix = matrix(1),
  vloadings = rep(0.7, 3),
  thresholds = matrix(rep(c(-0.5, 0.5), each = 3), nrow = 3),
  N = 100, R = 2, seed = 2026,
  file_dir = out_dir, file_prefix = "example"
)
result$avg_res_prob
```

simdata_grm

Simulate Ordinal Responses with a Probit Graded Response Model

Description

Simulates binary or ordinal item responses for unidimensional or multidimensional simple-structure models. Latent trait values are supplied by the user and remain fixed across replications.

Usage

```
simdata_grm(
  model,
  theta_matrix,
  a,
  d,
  N = NULL,
  R,
  method = c("U", "N"),
  file_dir = tempdir(),
  file_prefix = "sim",
  seed_num = 12345L
)
```

Arguments

model	A nonempty list. Each element contains the item indices loading on one factor. Every item from 1 through length(a) must occur once.
-------	---

<code>theta_matrix</code>	A numeric matrix with people in rows and factors in columns. A numeric vector is accepted for a one-factor model.
<code>a</code>	A numeric vector containing one discrimination parameter per item.
<code>d</code>	A numeric item-by-boundary matrix. For item j and boundary k , the cumulative probability is $P(Y[j] \geq k + 1 \mid \theta) = \text{pnorm}(a[j] * \theta - d[j, k])$. Values must increase across each row.
<code>N</code>	Optional sample size. When supplied, it must equal the number of rows in <code>theta_matrix</code> .
<code>R</code>	Positive integer number of replications.
<code>method</code>	Response-generation method. "U" uses independent $U(0, 1)$ draws. "N" uses independent standard-normal draws transformed to uniform probabilities. The methods have the same target distribution.
<code>file_dir</code>	Directory in which to write datasets and summaries. It is created recursively when necessary.
<code>file_prefix</code>	Filename prefix containing letters, digits, underscores, or hyphens.
<code>seed_num</code>	Positive integer base seed. Replication r uses <code>seed_num + r</code> .

Value

A list containing the model specification, population item parameters, response proportions by replication and item, and response proportions averaged over replications and items.

References

Samejima F (1969). "Estimation of latent ability using a response pattern of graded scores." *Psychometrika*, **34**(S1), 1–97. ISSN 0033-3123, doi:[10.1007/BF03372160](https://doi.org/10.1007/BF03372160).

Examples

```
theta <- matrix(seq(-2, 2, length.out = 100), ncol = 1)
out_dir <- tempfile("simPsyStudy-grm-")
result <- simdata_grm(
  model = list(1:3), theta_matrix = theta,
  a = rep(1, 3), d = matrix(rep(c(-0.5, 0.5), each = 3), nrow = 3),
  R = 2, method = "U", file_dir = out_dir,
  file_prefix = "example", seed_num = 100
)
result$avg_res_prob
```

`simdata_grm2`

Compatibility Wrapper for simdata_grm()

Description

`simdata_grm2()` is retained for scripts written against earlier development versions. New code should call `simdata_grm()`.

Usage

```
simdata_grm2(
  model,
  theta_matrix,
  a,
  d,
  R,
  method = c("U", "N"),
  file_dir = tempdir(),
  file_prefix = "sim",
  seed_num = 12345L
)
```

Arguments

model	A nonempty list. Each element contains the item indices loading on one factor. Every item from 1 through length(a) must occur once.
theta_matrix	A numeric matrix with people in rows and factors in columns. A numeric vector is accepted for a one-factor model.
a	A numeric vector containing one discrimination parameter per item.
d	A numeric item-by-boundary matrix. For item j and boundary k , the cumulative probability is $P(Y[j] \geq k + 1 \mid \theta) = \text{pnorm}(a[j] * \theta - d[j, k])$. Values must increase across each row.
R	Positive integer number of replications.
method	Response-generation method. "U" uses independent $U(0, 1)$ draws. "N" uses independent standard-normal draws transformed to uniform probabilities. The methods have the same target distribution.
file_dir	Directory in which to write datasets and summaries. It is created recursively when necessary.
file_prefix	Filename prefix containing letters, digits, underscores, or hyphens.
seed_num	Positive integer base seed. Replication r uses $\text{seed_num} + r$.

Value

The value returned by `simdata_grm()`.

simulate_condition	<i>Simulate a Full Factorial Set of Graded Response Conditions</i>
--------------------	--

Description

Generate a separate folder of replicated datasets for every combination of named theta, loading, and threshold sets using `simdata_grm()`.

Usage

```

simulate_condition(
  model,
  theta_sets,
  loading_sets,
  threshold_sets,
  replications = 5L,
  output_dir = tempfile("simPsyStudy-simulation-"),
  seed = 1234L,
  file_prefix = "bk",
  method = c("U", "N")
)

simulate_conditions(
  model,
  theta_sets,
  loading_sets,
  threshold_sets,
  replications = 5L,
  output_dir = tempfile("simPsyStudy-simulation-"),
  seed = 1234L,
  file_prefix = "bk",
  method = c("U", "N")
)

```

Arguments

model	A nonempty list of factors containing item indices. Every item from 1 to the total item count must appear exactly once.
theta_sets	Named list of finite numeric matrices. Rows are persons; columns correspond to factors in model. Theta values stay fixed across replications. Each column must have nonzero sample variance.
loading_sets	Named list of numeric vectors of standardized FAOV loadings, in item order, strictly between -1 and 1.
threshold_sets	Named list of increasing FAOV threshold vectors (shared by all items), or matrices with items in rows and increasing thresholds across columns. These are not GRM intercepts.
replications	Positive integer number of datasets per condition.
output_dir	Destination directory. Defaults to a new temporary path. Supply a persistent path to retain results. Existing run metadata or condition folders cause an error; files are never deliberately overwritten.
seed	Positive integer base seed. Condition i , replication r uses $\text{seed} + (i - 1) * \text{replications} + r$. The caller's random state is restored.
file_prefix	Filename prefix containing letters, digits, underscores, or hyphens. Defaults to "bk".

method Response-generation method passed to `simdata_grm()`. "U" uses independent uniform draws; "N" uses independent standard-normal draws transformed to uniform probabilities.

Details

Set names must be unique ignoring case and contain only letters, digits, underscores or hyphens, starting with a letter or digit. Threshold sets vary fastest, followed by loadings, then theta sets. Uses the probit graded response model in `simdata_grm()`, where $a = \text{loading} / \sqrt{1 - \text{loading}^2}$ and the boundary parameter is $d = \text{threshold} / \sqrt{1 - \text{loading}^2}$.

Each condition contains replication .dat files, study_cell.Rdata, parameters.rds, theta values and response summaries. The root contains conditions.csv, folders.Rdata and simulation_plan.rds (inputs, engine and session information). A failed condition stops the run, records its error in conditions.csv and preserves earlier outputs. There is no resume mode. `simulate_conditions()` is an alias with a plural name.

Value

Invisibly, a data frame with condition identifiers, input-set names, base seeds, sample sizes, replication counts, folder names, status and errors.

Examples

```
theta <- cbind(seq(-2, 2, length.out = 100),
               sin(seq(0, 2 * pi, length.out = 100)))
result <- simulate_condition(
  model = list(1:2, 3:4),
  theta_sets = list(example = theta),
  loading_sets = list(low = rep(0.4, 4), high = rep(0.7, 4)),
  threshold_sets = list(binary = 0),
  replications = 1, output_dir = tempfile("simPsyStudy-example-"), seed = 123)
result$status
```

to_cormatrix

Convert a Correlation Vector to a Correlation Matrix

Description

Convert a Correlation Vector to a Correlation Matrix

Usage

```
to_cormatrix(cor_vec, n_dim)
```

Arguments

cor_vec Pairwise correlations ordered as (1, 2), (1, 3), ..., (n_dim - 1, n_dim)'.
n_dim Positive integer number of dimensions.

Value

An n_dim by n_dim correlation matrix.

Examples

```
to_cormatrix(c(0.3, 0.4, 0.5), n_dim = 3)
```

to_covmatrix	<i>Convert a Correlation Matrix to a Covariance Matrix</i>
--------------	--

Description

Convert a Correlation Matrix to a Covariance Matrix

Usage

```
to_covmatrix(cor_matrix, sd_vec)
```

Arguments

- cor_matrix A finite symmetric correlation matrix.
- sd_vec A positive standard deviation for every matrix dimension.

Value

A covariance matrix.

Examples

```
correlation <- to_cormatrix(c(0.3, 0.5, 0.6), n_dim = 3)
to_covmatrix(correlation, sd_vec = c(0.95, 0.98, 1.1))
```

to_loadingmatrix	<i>Convert a Loading Vector to a Loading Matrix</i>
------------------	---

Description

Creates a simple-structure loading matrix with items in rows and factors in columns.

Usage

```
to_loadingmatrix(loading_vec, model)
```

Arguments

- loading_vec Finite numeric vector with one loading per item.
- model A list assigning every item index to exactly one factor.

Value

A numeric item-by-factor loading matrix.

Examples

```
to_loadingmatrix(rep(0.7, 6), list(c(1, 3, 5), c(2, 4, 6)))
```

TSK	<i>Thresholds, Skewness, and Kurtosis for Ordinal Responses</i>
-----	---

Description

Computes normal-theory thresholds and distribution summaries from a vector of ordinal response probabilities.

Usage

```
TSK(n, res_prop)
```

Arguments

- n Integer sample size greater than 3.
- res_prop Positive category probabilities that sum to 1.

Value

A list containing category and cumulative probabilities, thresholds, skewness, adjusted skewness, and three kurtosis measures.

Examples

```
TSK(n = 1200, res_prop = c(4, 16, 50, 50) / 120)
```

Index

`create_theta_mvn`, [2](#)

`diag_residuals`, [3](#)

`gamma_from_normal`, [3](#)
`gamma_from_normal()`, [5](#)
`gamma_from_normal2`, [5](#)
`gamma_from_normal2()`, [4](#)

`simdata_faov`, [6](#)
`simdata_grm`, [7](#)
`simdata_grm()`, [8](#), [9](#), [11](#)
`simdata_grm2`, [8](#)
`simulate_condition`, [9](#)
`simulate_conditions`
 (`simulate_condition`), [9](#)
`stats::dgamma()`, [4](#)
`stats::dnorm()`, [4](#)

`to_cormatrix`, [11](#)
`to_cormatrix()`, [2](#)
`to_covmatrix`, [12](#)
`to_covmatrix()`, [2](#)
`to_loadingmatrix`, [12](#)
`TSK`, [13](#)