Package 'dcm2'

May 8, 2026

Type Package

Title Calculating the M2 Model Fit Statistic for Diagnostic Classification Models

Version 1.0.2

Description A collection of functions for calculating the M2 model fit statistic for diagnostic classification models as described by Liu et al. (2016) <DOI:10.3102/1076998615621293>. These functions provide multiple sources of information for model fit according to the M2 statistic, including the M2 statistic, the *p* value for that M2 statistic, and the Root Mean Square Error of Approximation based on the M2 statistic.

License GPL-3

URL https://github.com/atlas-aai/dcm2

BugReports https://github.com/atlas-aai/dcm2/issues

Depends R (>= 3.6)

Imports dplyr (>= 0.8.4), glue (>= 1.4.2), magrittr (>= 1.5), methods (>= 4.1.0), modelr (>= 0.1.8), purrr (>= 0.3.3), Rcpp, rlang (>= 0.4.11), stringr (>= 1.4.0), tibble (>= 2.1.3), tidyr (>= 1.1.4)

Suggests covr, GDINA, roxygen2, spelling, testthat (>= 2.1.0)

LinkingTo Rcpp, RcppArmadillo (>= 0.9.800.1.0)

Config/testthat/edition 3

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

Language en-US

NeedsCompilation yes

Author Jeffrey Hoover [aut, cre, cph] (ORCID: <https://orcid.org/0000-0002-0276-0308>), W. Jake Thompson [aut] (ORCID: <https://orcid.org/0000-0001-7339-0300>), Wenchao Ma [ctb] (Author of Mord.cpp)

Maintainer Jeffrey Hoover <jeffrey.c.hoover@gmail.com>

Repository CRAN

Date/Publication 2023-03-21 19:50:02 UTC

Contents

`as_binary` *Make Binary Profiles*

Description

Given a number of attributes, `as_binary` will create all possible binary mastery profiles.

Usage

```
as_binary(x)
```

Arguments

`x`	The number of attributes

Value

A `2 ^ x` by `x` matrix

Examples

```
as_binary(3)
as_binary(4)
```

`calc_m2` *Calculate the M2*

Description

Calculate the M2

Usage

```
calc_m2(
  data,
  struc_params,
  pi_matrix,
  qmatrix,
  ci = 0.9,
  link = "logit",
 model_type = c("LCDM", "GDINA", "ACDM", "LLM", "RRUM", "DINO", "DINA", "BUGDINO")
)
```

Arguments

`data`	A data frame containing the raw data, where there is one row per respondent and one column per item
`struc_params`	A vector containing the structural parameters of the estimated model
`pi_matrix`	An item-by-class matrix containing the probability of a correct response by members of each latent class
`qmatrix`	A data frame containing the Q-matrix
`ci`	The confidence interval for the RMSEA, computed from the M2
`link`	A character containing the link function.
`model_type`	A character containing the model type (e.g., `LCDM`) that was estimated.

Value

A data frame containing:

- `m2`: The M2 statistic
- `df`: Degrees of freedom for the M2 statistic
- `pval`: *p*-value for the M2 statistic
- `rmsea`: Root mean square error of approximation
- `ci_lower`: Lower end of `ci` interval for RMSEA
- `ci_upper`: Upper end of `ci` interval for RMSEA
- `srmsr`: Standardized root mean square residual

Examples

```
possible_prof <- dcm2::as_binary(ncol(sample_data$q_matrix))

fit_dat <- sample_data$data %>%
             tidyr::pivot_wider(names_from = "item_id",
                                values_from = "score") %>%
             dplyr::select(-"resp_id") %>%
             as.matrix() %>%
             unname()
gdina_mod <- GDINA::GDINA(dat = fit_dat,
                          Q = data.frame(sample_data$q_matrix),
                          model = "logitGDINA",
                          control = list(conv.type = "neg2LL"))
struc_params <- gdina_mod$struc.parm
pi_matrix <- gdina_mod$LC.prob %>%
               as.matrix() %>%
               unname()
calc_m2(data = fit_dat, struc_params, pi_matrix,
        qmatrix = data.frame(sample_data$q_matrix), ci = 0.9, link = "logit",
        model_type = "LCDM")
```

`data_att1` *Simulated Data for a Single Attribute Assessment*

Description

A list containing data from a randomly simulated single-attribute assessment.

Usage

```
data_att1
```

Format

A list frame containing 4 `tibble` objects:

- `resp_profiles`: A `tibble` with 1000 rows and 2 columns. The first column indicates `resp_id` (i.e., the respondent identification number) and the second column indicates `att_1` (i.e., a binary indicator for whether the respondent mastered the first attribute).
- `q_matrix`: A `tibble` with 2 rows and 1 column. Each row corresponds to an assessment item, and the column entries provide a binary indicator for whether the item assessed the attribute.
- `item_params`: A `tibble` with 2 rows and 3 columns. Each row corresponds to an item. The first column indicates `item_id` (i.e., the item identification number). The second column indicates `intercept` (i.e., the true item intercept parameter for the item). The third column indicates `att_1` (i.e., the true item main effect parameter for the item).

- `data`: A `tibble` with 2000 rows and 3 columns. The first column indicates `resp_id` (i.e., the respondent identification number). The second column indicates `item_id` (i.e., the item identification number). The third column indicates `score` (i.e., the dichotomously scored item response).

`fit_m2` *Model Fit M2 Calculations*

Description

Estimate the M2 statistic as described by Liu et al. (2016).

Usage

```
fit_m2(model, ci = 0.9, ...)
```

Arguments

`model`	An estimated diagnostic classification model.
`ci`	The confidence interval for the RMSEA.
`...`	Unused, for extensibility.

Value

A data frame containing:

- `m2`: The M2 statistic
- `df`: Degrees of freedom for the M2 statistic
- `pval`: *p*-value for the M2 statistic
- `rmsea`: Root mean square error of approximation
- `ci_lower`: Lower end of `ci` interval for RMSEA
- `ci_upper`: Upper end of `ci` interval for RMSEA
- `srmsr`: Standardized root mean square residual

References

Liu, Y., Tian, W., & Xin, T. (2016). An application of M_2 statistic to evaluate the fit of cognitive diagnostic models. *Journal of Educational and Behavioral Statistics, 41*, 3-26. doi: 10.3102/1076998615621293

Examples

```
possible_prof <- dcm2::as_binary(ncol(sample_data$q_matrix))

fit_dat <- sample_data$data %>%
             tidyr::pivot_wider(names_from = "item_id",
                                values_from = "score") %>%
             dplyr::select(-"resp_id") %>%
             as.matrix() %>%
             unname()
gdina_mod <- GDINA::GDINA(dat = fit_dat,
                          Q = data.frame(sample_data$q_matrix),
                          model = "logitGDINA",
                          control = list(conv.type = "neg2LL"))
fit_m2(gdina_mod, ci = 0.9)
```

---

log_odds *Log-odds Transformation*

Description

These functions implement the log-odds (or logit) transformation. This is a common transformation for psychometric models that is used to put probabilities on a continuous scale.

Usage

```
logit(x)

inv_logit(x)
```

Arguments

x	A number to be transformed

Value

A transformed double

Examples

```
logit(0.6)
logit(0.5)

inv_logit(3.5)
inv_logit(0)
```

sample_data *Simulated Data for Testing Functions*

Description

A matrix with randomly simulated data to test the package functions.

Usage

```
sample_data
```

Format

A list frame containing 4 `tibble` objects:

- `resp_profiles`: A `tibble` with 1000 rows and 3 columns. The first column indicates `resp_id` (i.e., the respondent identification number). The second column indicates `att_1` (i.e., a binary indicator for whether the respondent mastered the first attribute). The third column indicates `att_2` (i.e., a binary indicator for whether the respondent mastered the second attribute).
- `q_matrix`: A `tibble` with 8 rows and 2 columns. Each row corresponds to an assessment item, and the column entries provide a binary indicator for whether the item assessed each of the attribute.
- `item_params`: A `tibble` with 8 rows and 5 columns. Each row corresponds to an item. The first column indicates `item_id` (i.e., the item identification number). The second column indicates `intercept` (i.e., the true item intercept parameter for the item). The third column indicates `att_1` (i.e., the true item main effect parameter for the first attribute for the item). The fourth column indicates `att_2` (i.e., the true item main effect parameter for the second attribute for the item). The fifth column indicates `att_1__att_2` (i.e., the true item interaction effect parameter for the first and second attributes).
- `data`: A `tibble` with 8000 rows and 3 columns. The first column indicates `resp_id` (i.e., the respondent identification number). The second column indicates `item_id` (i.e., the item identification number). The third column indicates `score` (i.e., the dichotomously scored item response).

Index