Package 'MAIHDA'

April 3, 2026

Type Package

Title Multilevel Analysis of Individual Heterogeneity and Discriminatory Accuracy

Version 0.1.0

Description Provides a comprehensive toolkit for conducting Multilevel Analysis of Individual Heterogeneity and Discriminatory Accuracy (MAIHDA). Methods are described in Merlo (2018) <doi:10.1016/j.socscimed.2017.12.018> and Evans et al. (2018) <doi:10.1016/j.socscimed.2017.11.011>. Automatically generates intersectional strata, fits analytical models, extracts statistics, and produces visualizations.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Depends R (>= 4.1.0)

Imports lme4 (>= 1.1-27), ggplot2 (>= 3.3.0), dplyr (>= 1.0.0), tidyr (>= 1.1.0), stats, rlang (>= 0.4.0)

Suggests brms (>= 2.15.0), testthat (>= 3.0.0), knitr, rmarkdown, boot (>= 1.3-20)

RoxygenNote 7.3.3

VignetteBuilder knitr

URL `https://github.com/hdbt/MAIHDA`, `https://hdbt.github.io/MAIHDA/`

BugReports `https://github.com/hdbt/MAIHDA/issues`

NeedsCompilation no

Author Hamid Bulut [aut, cre]

Maintainer Hamid Bulut <`hamid.bulut@ymail.com`>

Repository CRAN

Date/Publication 2026-04-03 08:10:02 UTC

Contents

`calculate_pvc` *Calculate Proportional Change in Between-Stratum Variance (PVC)*

Description

Calculates the proportional change in between-stratum variance (PVC) between two MAIHDA models. The PVC measures how much the between-stratum variance changes when moving from one model to another, and is calculated as: PVC = (Var_model1 - Var_model2) / Var_model1

Usage

```
calculate_pvc(
  model1,
  model2,
  bootstrap = FALSE,
  n_boot = 1000,
  conf_level = 0.95
)
```

Arguments

`model1`	A maihda_model object from `fit_maihda()`. This is the reference model (typically a simpler or baseline model).
`model2`	A maihda_model object from `fit_maihda()`. This is the comparison model (typically a more complex model with additional predictors).
`bootstrap`	Logical indicating whether to compute bootstrap confidence intervals for PVC. Default is FALSE.
`n_boot`	Number of bootstrap samples if bootstrap = TRUE. Default is 1000.
`conf_level`	Confidence level for bootstrap intervals. Default is 0.95.

Details

The PVC is interpreted as the proportional reduction (or increase if negative) in between-stratum variance when moving from model1 to model2. A positive PVC indicates that model2 explains some of the between-stratum variance present in model1, while a negative PVC suggests that model2 has more unexplained between-stratum variance.

When bootstrap = TRUE, the function resamples the data with replacement and refits both models for each bootstrap sample to obtain confidence intervals for the PVC estimate.

Value

A list containing:

`pvc`	The estimated proportional change in variance
`var_model1`	Between-stratum variance from model1
`var_model2`	Between-stratum variance from model2
`ci_lower`	Lower bound of confidence interval (if bootstrap = TRUE)
`ci_upper`	Upper bound of confidence interval (if bootstrap = TRUE)
`bootstrap`	Logical indicating if bootstrap was used

Examples

```
# Create strata and fit two models
strata_result <- make_strata(maihda_sim_data, c("gender", "race"))
model1 <- fit_maihda(health_outcome ~ age + (1 | stratum), data = strata_result$data)
model2 <- fit_maihda(health_outcome ~ age + gender + (1 | stratum), data = strata_result$data)

# Calculate PVC without bootstrap
pvc_result <- calculate_pvc(model1, model2)
print(pvc_result$pvc)

# Calculate PVC with bootstrap CI
# pvc_boot <- calculate_pvc(model1, model2, bootstrap = TRUE, n_boot = 500)
# print(pvc_boot)
```

`compare_maihda` *Compare MAIHDA Models*

Description

Compares variance partition coefficients (VPC/ICC) across multiple MAIHDA models, with optional bootstrap confidence intervals.

Usage

```
compare_maihda(
  ...,
  model_names = NULL,
  bootstrap = FALSE,
  n_boot = 1000,
  conf_level = 0.95
)
```

Arguments

`...`	Multiple maihda_model objects to compare.
`model_names`	Optional character vector of names for the models.
`bootstrap`	Logical indicating whether to compute bootstrap confidence intervals. Default is FALSE.
`n_boot`	Number of bootstrap samples if bootstrap = TRUE. Default is 1000.
`conf_level`	Confidence level for bootstrap intervals. Default is 0.95.

Value

A data frame comparing VPC/ICC across models with optional confidence intervals.

Examples

```
# Create strata and models using simulated data
strata_1 <- make_strata(maihda_sim_data, vars = c("gender", "race"))
strata_2 <- make_strata(maihda_sim_data, vars = c("gender", "race", "education"))

model1 <- fit_maihda(health_outcome ~ age + (1 | stratum), data = strata_1$data)
model2 <- fit_maihda(health_outcome ~ age + gender + (1 | stratum), data = strata_2$data)

# Compare without bootstrap
comparison <- compare_maihda(model1, model2,
                             model_names = c("Base", "With Gender"))

# Compare with bootstrap CI
comparison_boot <- compare_maihda(model1, model2,
                                  model_names = c("Base", "With Gender"),
                                  bootstrap = TRUE, n_boot = 500)
```

`fit_maihda` *Fit MAIHDA Model*

Description

Fits a multilevel model for MAIHDA (Multilevel Analysis of Individual Heterogeneity and Discriminatory Accuracy) using either lme4 or brms.

Usage

```
fit_maihda(formula, data, engine = "lme4", family = "gaussian", ...)
```

Arguments

`formula`	A formula specifying the model. Should include random effect for stratum (e.g., `outcome ~ fixed_vars + (1 | stratum)`).
`data`	A data frame containing the variables in the formula.
`engine`	Character string specifying which engine to use: "lme4" (default) or "brms".
`family`	Character string or family object specifying the model family. Common options: "gaussian", "binomial", "poisson". Default is "gaussian".
`...`	Additional arguments passed to `lmer`/`glmer` (lme4) or `brm` (brms).

Value

A maihda_model object containing:

`model`	The fitted model object (lme4 or brms)
`engine`	The engine used ("lme4" or "brms")
`formula`	The model formula
`data`	The data used for fitting
`family`	The family used
`strata_info`	The strata information from make_strata() if available, NULL otherwise

Examples

```
# Create strata
strata_result <- make_strata(maihda_sim_data, vars = c("gender", "race", "education"))

# Fit model with lme4
model <- fit_maihda(health_outcome ~ age + (1 | stratum),
                    data = strata_result$data,
                    engine = "lme4")

# Fit model with brms (if brms is available)
# model_brms <- fit_maihda(health_outcome ~ age + (1 | stratum),
#                          data = strata_result$data,
```

```
#                           engine = "brms")
```

---

maihda_sim_data *Simulated Health Data for MAIHDA Examples*

Description

A simulated dataset containing health outcomes and demographic variables for 500 individuals. This dataset is designed to demonstrate intersectional health inequalities suitable for MAIHDA analysis. The data includes main effects and intersectional effects between gender, race, and education.

Usage

```
data("maihda_sim_data")
```

Format

A data frame with 500 observations on the following 6 variables:

`id` Integer identifier for each individual

`gender` Character variable indicating gender ("Male" or "Female")

`race` Character variable indicating race/ethnicity ("White", "Black", "Hispanic", or "Asian")

`education` Character variable indicating education level ("High School", "Some College", "College", or "Graduate")

`age` Numeric variable for age in years (range: 18-80)

`health_outcome` Numeric variable representing a health score (higher is better)

Details

The health outcome was simulated with:

- Main effects of gender, race, education, and age
- Intersectional effects (e.g., Black women, men with high school education)
- Random noise with standard deviation of 10

The data demonstrates typical patterns in health inequalities research where outcomes vary both by individual characteristics and their intersections.

Examples

```
data(maihda_sim_data)

# View structure
str(maihda_sim_data)

# Summary statistics
summary(maihda_sim_data)


# Example MAIHDA analysis
library(MAIHDA)

# Create strata
strata_result <- make_strata(maihda_sim_data, vars = c("gender", "race"))

# Fit model
model <- fit_maihda(health_outcome ~ age + (1 | stratum),
                    data = strata_result$data)

# Summarize
summary_maihda(model)

# Visualize
plot_maihda(model, type = "caterpillar")
```

`make_strata` *Create Strata from Multiple Variables*

Description

This function creates strata (intersectional categories) from multiple categorical variables in a dataset.

Usage

```
make_strata(data, vars, sep = "_", min_n = 1)
```

Arguments

`data`	A data frame containing the variables to create strata from.
`vars`	Character vector of variable names to use for creating strata.
`sep`	Separator to use between variable values when creating stratum labels. Default is "_".
`min_n`	Minimum number of observations required for a stratum to be included. Strata with fewer observations will be coded as NA. Default is 1.

Details

If any of the specified variables has a missing value (NA) for a given observation, that observation will be assigned to the NA stratum (stratum = NA), rather than creating a stratum that includes the missing value.

The strata_info data frame is also attached as an attribute to the data, which allows fit_maihda() to automatically capture stratum labels for use in plots and summaries.

Value

A list with two elements:

`data`	The original data frame with an added 'stratum' column. The strata_info is also attached as an attribute for use by fit_maihda()
`strata_info`	A data frame with information about each stratum including counts and the combination of variable values

Examples

```
# Create strata from gender and race variables
result <- make_strata(maihda_sim_data, vars = c("gender", "race"))
print(result$strata_info)
```

`plot_comparison` *Plot Model Comparison*

Description

Creates a plot comparing VPC/ICC across multiple models.

Usage

```
plot_comparison(comparison_df)
```

Arguments

`comparison_df`	A data frame from `compare_maihda()`.

Value

A ggplot2 object.

Examples

```
# Create strata and models using simulated data
strata_1 <- make_strata(maihda_sim_data, vars = c("gender", "race"))
strata_2 <- make_strata(maihda_sim_data, vars = c("gender", "race", "education"))

model1 <- fit_maihda(health_outcome ~ age + (1 | stratum), data = strata_1$data)
model2 <- fit_maihda(health_outcome ~ age + gender + (1 | stratum), data = strata_2$data)

comparison <- compare_maihda(model1, model2, bootstrap = TRUE)
plot_comparison(comparison)
```

`plot_maihda` *Plot MAIHDA Model Results*

Description

Creates various plots for visualizing MAIHDA model results including caterpillar plots, variance partition coefficient comparisons, observed vs. shrunken estimates, and predicted subgroup values with confidence intervals.

Usage

```
plot_maihda(
  object,
  type = c("caterpillar", "vpc", "obs_vs_shrunken", "predicted"),
  summary_obj = NULL,
  n_strata = 50,
  ...
)
```

Arguments

`object`	A maihda_model object from `fit_maihda()`.
`type`	Character string specifying plot type: • "caterpillar": Caterpillar plot of stratum random effects • "vpc": Variance partition coefficient visualization • "obs_vs_shrunken": Observed vs. shrunken stratum means • "predicted": Predicted values for each stratum with confidence intervals
`summary_obj`	Optional maihda_summary object from `summary_maihda()`. If NULL, will be computed.
`n_strata`	Maximum number of strata to display in caterpillar plot or predicted plot. Default is 50. Use NULL for all strata.
`...`	Additional arguments (not currently used).

Value

A ggplot2 object.

Examples

```
strata_result <- make_strata(maihda_sim_data, vars = c("gender", "race"))
model <- fit_maihda(health_outcome ~ age + (1 | stratum), data = strata_result$data)

# Caterpillar plot
plot_maihda(model, type = "caterpillar")

# VPC plot
plot_maihda(model, type = "vpc")

# Observed vs shrunken plot
plot_maihda(model, type = "obs_vs_shrunken")

# Predicted values with confidence intervals
plot_maihda(model, type = "predicted")
```

`predict_maihda` *Predict from MAIHDA Model*

Description

Makes predictions from a fitted MAIHDA model, either at the stratum level or individual level.

Usage

```
predict_maihda(object, newdata = NULL, type = c("individual", "strata"), ...)
```

Arguments

`object`	A maihda_model object from `fit_maihda()`.
`newdata`	Optional data frame for making predictions. If NULL, uses the original data from model fitting.
`type`	Character string specifying prediction type: • "individual": Individual-level predictions including random effects • "strata": Stratum-level predictions (random effects only)
`...`	Additional arguments passed to predict method of underlying model.

Value

Depending on type:

- For "individual": A numeric vector of predicted values
- For "strata": A data frame with stratum ID and predicted random effect

Examples

```
strata_result <- make_strata(maihda_sim_data, vars = c("gender", "race"))
model <- fit_maihda(health_outcome ~ age + (1 | stratum), data = strata_result$data)

# Individual predictions
pred_ind <- predict_maihda(model, type = "individual")

# Stratum predictions
pred_strata <- predict_maihda(model, type = "strata")
```

`print.maihda_model` *Print method for maihda_model*

Description

Print method for maihda_model

Usage

```
## S3 method for class 'maihda_model'
print(x, ...)
```

Arguments

`x`	A maihda_model object
`...`	Additional arguments

Value

No return value, called for side effects.

`print.maihda_strata` *Print method for maihda_strata objects*

Description

Print method for maihda_strata objects

Usage

```
## S3 method for class 'maihda_strata'
print(x, ...)
```

Arguments

`x`	A maihda_strata object
`...`	Additional arguments (not used)

Value

No return value, called for side effects.

`print.maihda_summary` *Print method for maihda_summary objects*

Description

Print method for maihda_summary objects

Usage

```
## S3 method for class 'maihda_summary'
print(x, ...)
```

Arguments

`x`	A maihda_summary object
`...`	Additional arguments (not used)

Value

No return value, called for side effects.

`print.pvc_result` *Print method for PVC results*

Description

Print method for PVC results

Usage

```
## S3 method for class 'pvc_result'
print(x, ...)
```

Arguments

`x`	A pvc_result object
`...`	Additional arguments

Value

No return value, called for side effects.

`summary_maihda` *Summarize MAIHDA Model*

Description

Provides a summary of a MAIHDA model including variance partition coefficients (VPC/ICC) and stratum-specific estimates.

Usage

```
summary_maihda(
  object,
  bootstrap = FALSE,
  n_boot = 1000,
  conf_level = 0.95,
  ...
)
```

Arguments

`object`	A maihda_model object from `fit_maihda()`.
`bootstrap`	Logical indicating whether to compute bootstrap confidence intervals for VPC/ICC. Default is FALSE.
`n_boot`	Number of bootstrap samples if bootstrap = TRUE. Default is 1000.
`conf_level`	Confidence level for bootstrap intervals. Default is 0.95.
`...`	Additional arguments (not currently used).

Value

A maihda_summary object containing:

`vpc`	Variance Partition Coefficient (ICC) with optional CI
`variance_components`	Data frame of variance components
`stratum_estimates`	Data frame of stratum-specific random effects with labels if available
`fixed_effects`	Fixed effects estimates
`model_summary`	Original model summary

Examples

```
strata_result <- make_strata(maihda_sim_data, vars = c("gender", "race"))
model <- fit_maihda(health_outcome ~ age + (1 | stratum), data = strata_result$data)
summary_result <- summary_maihda(model)

# With bootstrap CI
# summary_boot <- summary_maihda(model, bootstrap = TRUE, n_boot = 50)
```

Index