Package 'ebdm'

July 17, 2025

Type Package

Title Estimating Bivariate Dependency from Marginal Data

Version 1.1.0

Description Provides maximum likelihood methods to estimate bivariate dependency (correlation) from marginal summary statistics in multi-study settings.
The package supports both binary and continuous variables assumed to follow a bivariate normal distribution,
enabling privacy-preserving joint estimation when individual-level data are unavailable.
The binary method is fully described in the manuscript by Shang, Tsao and Zhang (2025) <doi:10.48550/arXiv.2505.03995>: ``Estimating the Joint Distribution of Two Binary Variables from Their Marginal Summaries".

License GPL (>= 3)

Encoding UTF-8

LazyData true

Depends R (>= 3.5.0)

Imports stats

RoxygenNote 7.3.2

NeedsCompilation no

Author Longwen Shang [aut, cre],
Min Tsao [aut],
Xuekui Zhang [aut]

Maintainer Longwen Shang <shanglongwen0918@gmail.com>

Repository CRAN

Date/Publication 2025-07-17 19:40:01 UTC

Contents

bin_example *Example Dataset*

Description

Simulated dataset for testing the `cor_bin()` function.

Usage

```
data(bin_example)
```

Format

A data frame with 3 columns:

ni Sample size per study

xi Count of first binary variable

yi Count of second binary variable

cont_example *Example Data: Continuous Variables*

Description

Simulated dataset for testing the `cor_cont()` function.

Usage

```
data(cont_example)
```

Format

A data frame with 5 columns:

Sample_Size Sample size for each study.

Mean_X Sample mean of variable X.

Mean_Y Sample mean of variable Y.

Variance_X Sample variance of variable X.

Variance_Y Sample variance of variable Y.

`cor_bin` *Estimate the Joint Distribution of Two Binary Variables from Marginal Summaries*

Description

Performs maximum likelihood estimation (MLE) of the joint distribution of two binary variables using only marginal summary data from multiple studies.

Usage

```
cor_bin(ni, xi, yi, ci_method = c("none", "normal", "lr"))
```

Arguments

`ni`	Numeric vector. Sample sizes for each dataset.
`xi`	Numeric vector. Count of observations where variable 1 equals 1.
`yi`	Numeric vector. Count of observations where variable 2 equals 1.
`ci_method`	Character string. Method for confidence interval computation. Options are `"none"` (default), `"normal"`, or `"lr"` (likelihood ratio).

Value

A named list with point estimates, variance, standard error, and confidence interval (if requested).

p1_hat Estimated marginal probability for variable 1.

p2_hat Estimated marginal probability for variable 2.

p11_hat Estimated joint probability.

var_hat Estimated variance of `p11_hat`.

sd_hat Standard error of `p11_hat`.

ci Confidence interval for `p11_hat`, if requested.

Examples

```
data(bin_example)
cor_bin(bin_example$ni, bin_example$xi, bin_example$yi, ci_method = "lr")
```

cor_cont *Estimate the Bivariate Normal Distribution from Marginal Summaries*

Description

Estimate the correlation coefficient ρ (and marginal means / SDs) of two normally-distributed variables using summary-level data from multiple independent studies.

Usage

```
cor_cont(
  n,
  xbar,
  ybar,
  s2x = NULL,
  s2y = NULL,
  method = c("proposed", "weighted"),
  ci_method = c("none", "normal", "lr")
)
```

Arguments

n	Numeric vector. Sample size of each study.
xbar, ybar	Numeric vectors. Sample means of the two variables.
s2x, s2y	Numeric vectors. Sample variances; required for method = "proposed".
method	Character. "proposed" uses the proposed MLE method in the paper; "weighted" replicates the weighted mean based method (Baseline) when no variances are available.
ci_method	Confidence interval type: "none", "normal", or "lr" (likelihood ratio). Only implemented when method = "proposed".

Value

A list with elements

- mu_x, mu_y : estimated marginal means
- sigma_x, sigma_y : estimated SDs
- rho : estimated correlation
- se : standard error of rho (proposed only)
- ci : confidence interval for rho (if requested)

Examples

```
data(cont_example)
# Example with full summaries
cor_cont(cont_example$Sample_Size, cont_example$Mean_X, cont_example$Mean_Y,
 cont_example$Variance_X, cont_example$Variance_Y, method = "proposed", ci_method = "lr")

# Only means + n, weighted mean method
cor_cont(cont_example$Sample_Size, cont_example$Mean_X, cont_example$Mean_Y, method = "weighted")
```

Index