Package ‘multibias’

May 10, 2025

Type Package

Title Multiple Bias Analysis in Causal Inference

Version 1.7.1

Maintainer Paul Brendel <pcbrendel@gmail.com>

Description Quantify the causal effect of a binary exposure on a binary outcome with adjustment for multiple biases. The functions can simultaneously adjust for any combination of uncontrolled confounding, exposure/outcome misclassification, and selection bias.
The underlying method generalizes the concept of combining inverse probability of selection weighting with predictive value weighting. Simultaneous multi-bias analysis can be used to enhance the validity and transparency of real-world evidence obtained from observational, longitudinal studies. Based on the work from Paul Brendel, Aracelis Torres, and Onyebuchi Arah (2023) <doi:10.1093/ije/dyad001>.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Depends R (>= 2.10)

RoxygenNote 7.3.2

Imports dplyr (>= 1.1.3), lifecycle (>= 1.0.3), magrittr (>= 2.0.3), rlang (>= 1.1.1), broom (>= 1.0.5)

Suggests knitr, rmarkdown, MASS, testthat (>= 3.0.0)

URL https://github.com/pcbrendel/multibias,
http://www.paulbrendel.com/multibias/

BugReports https://github.com/pcbrendel/multibias/issues

Config/testthat/edition 3

VignetteBuilder knitr

NeedsCompilation no

Author Paul Brendel [aut, cre, cph]

Repository CRAN

Date/Publication 2025-05-10 18:40:06 UTC

Contents

`bias_params` *Represent bias parameters*

Description

`bias_params` is one of two different options to represent bias assumptions for bias adjustment. The `multibias_adjust()` function will apply the assumptions from these models and use them to adjust for biases in the observed data. It takes one input, a list, where each item in the list corresponds to the necessary models for bias adjustment. See below for bias models.

For each of the following bias models, the variables are defined:

- X = True exposure
- X* = Misclassified exposure
- Y = True outcome

- Y* = Misclassified outcome
- C = Known confounder
- j = Number of known confounders
- U = Uncontrolled confounder
- S = Selection indicator

Uncontrolled confounding $logit(P(U=1)) = \alpha_0 + \alpha_1 X + \alpha_2 Y + \alpha_{2+j} C_j$

Exposure misclassification $logit(P(X=1)) = \delta_0 + \delta_1 X^* + \delta_2 Y + \delta_{2+j} C_j$

Outcome misclassification $logit(P(Y=1)) = \delta_0 + \delta_1 X + \delta_2 Y^* + \delta_{2+j} C_j$

Selection bias $logit(P(S=1)) = \beta_0 + \beta_1 X + \beta_2 Y$

Uncontrolled Confounding & Exposure Misclassification (Option 1) $logit(P(U=1)) = \alpha_0 + \alpha_1 X + \alpha_2 Y$
$logit(P(X=1)) = \delta_0 + \delta_1 X^* + \delta_2 Y + \delta_{2+j} C_j$

Uncontrolled Confounding & Exposure Misclassification (Option 2) $log(P(X=1,U=0)/P(X=0,U=0)) = \gamma_{1,0} + \gamma_{1,1} X^* + \gamma_{1,2} Y + \gamma_{1,2+j} C_j$
$log(P(X=0,U=1)/P(X=0,U=0)) = \gamma_{2,0} + \gamma_{2,1} X^* + \gamma_{2,2} Y + \gamma_{2,2+j} C_j$
$log(P(X=1,U=1)/P(X=0,U=0)) = \gamma_{3,0} + \gamma_{3,1} X^* + \gamma_{3,2} Y + \gamma_{3,2+j} C_j$

Uncontrolled Confounding & Outcome Misclassification (Option 1) $logit(P(U=1)) = \alpha_0 + \alpha_1 X + \alpha_2 Y$
$logit(P(Y=1)) = \delta_0 + \delta_1 X + \delta_2 Y^* + \delta_{2+j} C_j$

Uncontrolled Confounding & Outcome Misclassification (Option 2) $log(P(U=1,Y=0)/P(U=0,Y=0)) = \gamma_{1,0} + \gamma_{1,1} X + \gamma_{1,2} Y^* + \gamma_{1,2+j} C_j$
$log(P(U=0,Y=1)/P(U=0,Y=0)) = \gamma_{2,0} + \gamma_{2,1} X + \gamma_{2,2} Y^* + \gamma_{2,2+j} C_j$
$log(P(U=1,Y=1)/P(U=0,Y=0)) = \gamma_{3,0} + \gamma_{3,1} X + \gamma_{3,2} Y^* + \gamma_{3,2+j} C_j$

Uncontrolled Confounding & Selection Bias $logit(P(U=1)) = \alpha_0 + \alpha_1 X + \alpha_2 Y + \alpha_{2+j} C_j$
$logit(P(S=1)) = \beta_0 + \beta_1 X + \beta_2 Y$

Exposure Misclassification & Outcome Misclassification (Option 1) $logit(P(X=1)) = \delta_0 + \delta_1 X^* + \delta_2 Y^* + \delta_{2+j} C_j$
$logit(P(Y=1)) = \beta_0 + \beta_1 X + \beta_2 Y^* + \beta_{2+j} C_j$

Exposure Misclassification & Outcome Misclassification (Option 2) $log(P(X=1,Y=0)/P(X=0,Y=0)) = \gamma_{1,0} + \gamma_{1,1} X^* + \gamma_{1,2} Y^* + \gamma_{1,2+j} C_j$
$log(P(X=0,Y=1)/P(X=0,Y=0)) = \gamma_{2,0} + \gamma_{2,1} X^* + \gamma_{2,2} Y^* + \gamma_{2,2+j} C_j$
$log(P(X=1,Y=1)/P(X=0,Y=0)) = \gamma_{3,0} + \gamma_{3,1} X^* + \gamma_{3,2} Y^* + \gamma_{3,2+j} C_j$

Exposure Misclassification & Selection Bias $logit(P(X=1)) = \delta_0 + \delta_1 X^* + \delta_2 Y + \delta_{2+j} C_j$
$logit(P(S=1)) = \beta_0 + \beta_1 X^* + \beta_2 Y + \beta_{2+j} C_j$

Outcome Misclassification & Selection Bias $logit(P(Y=1)) = \delta_0 + \delta_1 X + \delta_2 Y^* + \delta_{2+j} C_j$
$logit(P(S=1)) = \beta_0 + \beta_1 X + \beta_2 Y^* + \beta_{2+j} C_j$

Uncontrolled Confounding, Exposure Misclassification, and Selection Bias (Option 1) $logit(P(U=1)) = \alpha_0 + \alpha_1 X + \alpha_2 Y$
$logit(P(X=1)) = \delta_0 + \delta_1 X^* + \delta_2 Y + \delta_{2+j} C_j$
$logit(P(S=1)) = \beta_0 + \beta_1 X^* + \beta_2 Y + \beta_{2+j} C_j$

Uncontrolled Confounding, Exposure Misclassification, and Selection Bias (Option 2) $log(P(X = 1, U = 0)/P(X = 0, U = 0)) = \gamma_{1,0} + \gamma_{1,1}X^* + \gamma_{1,2}Y + \gamma_{1,2+j}C_j$
$log(P(X = 0, U = 1)/P(X = 0, U = 0)) = \gamma_{2,0} + \gamma_{2,1}X^* + \gamma_{2,2}Y + \gamma_{2,2+j}C_j$
$log(P(X = 1, U = 1)/P(X = 0, U = 0)) = \gamma_{3,0} + \gamma_{3,1}X^* + \gamma_{3,2}Y + \gamma_{3,2+j}C_j$
$logit(P(S = 1)) = \beta_0 + \beta_1 X^* + \beta_2 Y + \beta_{2+j}C_j$

Uncontrolled Confounding, Outcome Misclassification, and Selection Bias (Option 1) $logit(P(U = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 Y$
$logit(P(Y = 1)) = \delta_0 + \delta_1 X + \delta_2 Y^* + \delta_{2+j}C_j$
$logit(P(S = 1)) = \beta_0 + \beta_1 X + \beta_2 Y^* + \beta_{2+j}C_j$

Uncontrolled Confounding, Outcome Misclassification, and Selection Bias (Option 2) $log(P(U = 1, Y = 0)/P(U = 0, Y = 0)) = \gamma_{1,0} + \gamma_{1,1}X + \gamma_{1,2}Y^* + \gamma_{1,2+j}C_j$
$log(P(U = 0, Y = 1)/P(U = 0, Y = 0)) = \gamma_{2,0} + \gamma_{2,1}X + \gamma_{2,2}Y^* + \gamma_{2,2+j}C_j$
$log(P(U = 1, Y = 1)/P(U = 0, Y = 0)) = \gamma_{3,0} + \gamma_{3,1}X + \gamma_{3,2}Y^* + \gamma_{3,2+j}C_j$
$logit(P(S = 1)) = \beta_0 + \beta_1 X + \beta_2 Y^* + \beta_{2+j}C_j$

Usage

```
bias_params(coef_list)
```

Arguments

`coef_list`	List of coefficient values from the above options of models. Each item of the list is an equation. The left side of the equation identifies the model (i.e., "u" for the model predicting the uncontrolled confounder). For the multinomial models, specify the value here based on the numerator (i.e., "x1u0", "x0u1", "x1u1" for the three multinomial models in Uncontrolled Confounding & Exposure Misclassification, Option 2) The right side of the equation is the vector of values corresponding to the model coefficients (from left to right).

Examples

```
list_for_uc <- list(
  u = c(-0.19, 0.61, 0.70, -0.09, 0.10, -0.15)
)

bp_uc <- bias_params(coef_list = list_for_uc)

list_for_em_om <- list(
  x1y0 = c(-2.18, 1.63, 0.23, 0.36),
  x0y1 = c(-3.17, 0.22, 1.60, 0.40),
  x1y1 = c(-4.76, 1.82, 1.83, 0.72)
)

bp_em_om <- bias_params(coef_list = list_for_em_om)
```

data_observed — *Represent observed causal data*

Description

data_observed combines the observed dataframe with specific identification of the columns corresponding to the exposure, outcome, and confounders. It is an essential input of the multibias_adjust() function.

Usage

```
data_observed(data, bias, exposure, outcome, confounders = NULL)
```

Arguments

data	Dataframe for bias analysis.
bias	String type(s) of bias distorting the effect of the exposure on the outcome. Can choose from a subset of the following: "uc", "em", "om", "sel". These correspond to uncontrolled confounding, exposure misclassification, outcome misclassification, and selection bias, respectively.
exposure	String name of the column in data corresponding to the exposure variable.
outcome	String name of the column in data corresponding to the outcome variable.
confounders	String name(s) of the column(s) in data corresponding to the confounding variable(s).

Examples

```
df <- data_observed(
  data = df_sel,
  bias = "uc",
  exposure = "X",
  outcome = "Y",
  confounders = c("C1", "C2", "C3")
)
```

data_validation — *Represent validation causal data*

Description

data_validation is one of two different options to represent bias assumptions for bias adjustment. It combines the validation dataframe with specific identification of the appropriate columns for bias adjustment, including: true exposure, true outcome, confounders, misclassified exposure, misclassified outcome, and selection. The purpose of validation data is to use an external data source to transport the necessary causal relationships that are missing in the observed data.

Usage

```
data_validation(
  data,
  true_exposure,
  true_outcome,
  confounders = NULL,
  misclassified_exposure = NULL,
  misclassified_outcome = NULL,
  selection = NULL
)
```

Arguments

`data`	Dataframe of validation data
`true_exposure`	String name of the column in `data` corresponding to the true exposure.
`true_outcome`	String name of the column in `data` corresponding to the true outcome.
`confounders`	String name(s) of the column(s) in `data` corresponding to the confounding variable(s).
`misclassified_exposure`	String name of the column in `data` corresponding to the misclassified exposure.
`misclassified_outcome`	String name of the column in `data` corresponding to the misclassified outcome.
`selection`	String name of the column in `data` corresponding to the selection indicator.

Examples

```
df <- data_validation(
  data = df_sel_source,
  true_exposure = "X",
  true_outcome = "Y",
  confounders = c("C1", "C2", "C3"),
  selection = "S"
)
```

`df_em` *Simulated data with exposure misclassification*

Description

Data containing one source of bias, three known confounders, and 100,000 observations. This data is obtained from `df_emc_source` by removing the column *X*. The resulting data corresponds to what a researcher would see in the real-world: a misclassified exposure, *Xstar*, and no data on the true exposure. As seen in `df_emc_source`, the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_em
```

Format

A dataframe with 100,000 rows and 5 columns:

Xstar misclassified exposure, 1 = present and 0 = absent

Y outcome, 1 = present and 0 = absent

C1 1st confounder, 1 = present and 0 = absent

C2 2nd confounder, 1 = present and 0 = absent

C3 3rd confounder, 1 = present and 0 = absent

`df_em_om` *Simulated data with exposure misclassification and outcome misclassification*

Description

Data containing two sources of bias, three known confounders, and 100,000 observations. This data is obtained from `df_emc_omc_source` by removing the columns *X* and *Y*. The resulting data corresponds to what a researcher would see in the real-world: a misclassified exposure, *Xstar*, and a misclassified outcome, *Ystar*. As seen in `df_em_om_source`, the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_em_om
```

Format

A dataframe with 100,000 rows and 5 columns:

Xstar misclassified exposure, 1 = present and 0 = absent

Ystar misclassified outcome, 1 = present and 0 = absent

C1 1st confounder, 1 = present and 0 = absent

C2 2nd confounder, 1 = present and 0 = absent

C3 3rd confounder, 1 = present and 0 = absent

`df_em_om_source` *Data source for* `df_em_om`

Description

Data with complete information on the two sources of bias, three known confounders, and 100,000 observations. This data is used to derive `df_em_om` and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with `df_em_om`. With this source data, the fitted regression $logit(P(Y = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 C1 + \alpha_3 C2 + \alpha_4 C3$ shows that the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_em_om_source
```

Format

A dataframe with 100,000 rows and 7 columns:

X true exposure, 1 = present and 0 = absent

Y outcome, 1 = present and 0 = absent

C1 1st confounder, 1 = present and 0 = absent

C2 2nd confounder, 1 = present and 0 = absent

C3 3rd confounder, 1 = present and 0 = absent

Xstar misclassified exposure, 1 = present and 0 = absent

Ystar misclassified outcome, 1 = present and 0 = absent

`df_em_sel` *Simulated data with exposure misclassification and selection bias*

Description

Data containing two sources of bias, three known confounders, and 100,000 observations. This data is obtained by sampling with replacement with probability = *S* from `df_em_sel_source` then removing the columns *X* and *S*. The resulting data corresponds to what a researcher would see in the real-world: a misclassified exposure, *Xstar*, and missing data for those not selected into the study (*S*=0). As seen in `df_em_sel_source`, the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_em_sel
```

Format

A dataframe with 100,000 rows and 5 columns:

Xstar misclassified exposure, 1 = present and 0 = absent

Y outcome, 1 = present and 0 = absent

C1 1st confounder, 1 = present and 0 = absent

C2 2nd confounder, 1 = present and 0 = absent

C3 3rd confounder, 1 = present and 0 = absent

`df_em_sel_source` *Data source for* `df_em_sel`

Description

Data with complete information on the two sources of bias, three known confounders, and 100,000 observations. This data is used to derive `df_em_sel` and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with `df_em_sel`. With this source data, the fitted regression $logit(P(Y = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 C1 + \alpha_3 C2 + \alpha_4 C3$ shows that the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_em_sel_source
```

Format

A dataframe with 100,000 rows and 7 columns:

X true exposure, 1 = present and 0 = absent

Y outcome, 1 = present and 0 = absent

C1 1st confounder, 1 = present and 0 = absent

C2 2nd confounder, 1 = present and 0 = absent

C3 3rd confounder, 1 = present and 0 = absent

Xstar misclassified exposure, 1 = present and 0 = absent

S selection, 1 = selected into the study and 0 = not selected into the study

df_em_source *Data source for* df_em

Description

Data with complete information on one sources of bias, three known confounders, and 100,000 observations. This data is used to derive df_em and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with df_em. With this source data, the fitted regression $logit(P(Y=1)) = \alpha_0 + \alpha_1 X + \alpha_2 C1 + \alpha_3 C2 + \alpha_4 C3$ shows that the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_em_source
```

Format

A dataframe with 100,000 rows and 6 columns:

X exposure, 1 = present and 0 = absent

Y true outcome, 1 = present and 0 = absent

C1 1st confounder, 1 = present and 0 = absent

C2 2nd confounder, 1 = present and 0 = absent

C3 3rd confounder, 1 = present and 0 = absent

Xstar misclassified exposure, 1 = present and 0 = absent

---

df_om *Simulated data with outcome misclassification*

Description

Data containing one source of bias, three known confounders, and 100,000 observations. This data is obtained from df_om_source by removing the column *Y*. The resulting data corresponds to what a researcher would see in the real-world: a misclassified outcome, *Ystar*, and no data on the true outcome. As seen in df_om_source, the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_om
```

Format

A dataframe with 100,000 rows and 5 columns:

X exposure, 1 = present and 0 = absent

Ystar misclassified outcome, 1 = present and 0 = absent

C1 1st confounder, 1 = present and 0 = absent

C2 2nd confounder, 1 = present and 0 = absent

C3 3rd confounder, 1 = present and 0 = absent

`df_om_sel` *Simulated data with outcome misclassification and selection bias*

Description

Data containing two sources of bias, a known confounder, and 100,000 observations. This data is obtained by sampling with replacement with probability = *S* from `df_om_sel_source` then removing the columns *Y* and *S*. The resulting data corresponds to what a researcher would see in the real-world: a misclassified outcome, *Ystar*, and missing data for those not selected into the study (*S*=0). As seen in `df_om_sel_source`, the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_om_sel
```

Format

A dataframe with 100,000 rows and 5 columns:

X exposure, 1 = present and 0 = absent

Ystar misclassified outcome, 1 = present and 0 = absent

C1 1st confounder, 1 = present and 0 = absent

C2 2nd confounder, 1 = present and 0 = absent

C3 3rd confounder, 1 = present and 0 = absent

df_om_sel_source *Data source for* `df_om_sel`

Description

Data with complete information on the two sources of bias, a known confounder, and 100,000 observations. This data is used to derive `df_om_sel` and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with `df_om_sel`. With this source data, the fitted regression $logit(P(Y = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 C1 + \alpha_3 C2 + \alpha_4 C3$ shows that the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_om_sel_source
```

Format

A dataframe with 100,000 rows and 7 columns:

X exposure, 1 = present and 0 = absent

Y true outcome, 1 = present and 0 = absent

C1 1st confounder, 1 = present and 0 = absent

C2 2nd confounder, 1 = present and 0 = absent

C3 3rd confounder, 1 = present and 0 = absent

Ystar misclassified outcome, 1 = present and 0 = absent

S selection, 1 = selected into the study and 0 = not selected into the study

df_om_source *Data source for* `df_om`

Description

Data with complete information on one sources of bias, three known confounders, and 100,000 observations. This data is used to derive `df_om` and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with `df_om`. With this source data, the fitted regression $logit(P(Y = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 C1 + \alpha_3 C2 + \alpha_4 C3$ shows that the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_om_source
```

Format

A dataframe with 100,000 rows and 6 columns:

X exposure, 1 = present and 0 = absent

Y true outcome, 1 = present and 0 = absent

C1 1st confounder, 1 = present and 0 = absent

C2 2nd confounder, 1 = present and 0 = absent

C3 3rd confounder, 1 = present and 0 = absent

Ystar misclassified outcome, 1 = present and 0 = absent

`df_sel` *Simulated data with selection bias*

Description

Data containing one source of bias, three known confounders, and 100,000 observations. This data is obtained by sampling with replacement with probability = *S* from `df_sel_source` then removing the *S* column. The resulting data corresponds to what a researcher would see in the real-world: missing data for those not selected into the study (*S*=0). As seen in `df_sel_source`, the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_sel
```

Format

A dataframe with 100,000 rows and 5 columns:

X exposure, 1 = present and 0 = absent

Y outcome, 1 = present and 0 = absent

C1 1st confounder, 1 = present and 0 = absent

C2 2nd confounder, 1 = present and 0 = absent

C3 3rd confounder, 1 = present and 0 = absent

df_sel_source *Data source for* df_sel

Description

Data with complete information on study selection, three known confounders, and 100,000 observations. This data is used to derive df_sel and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with df_sel. With this source data, the fitted regression $logit(P(Y = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 C1 + \alpha_3 C2 + \alpha_4 C3$ shows that the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_sel_source
```

Format

A dataframe with 100,000 rows and 6 columns:

X true exposure, 1 = present and 0 = absent

Y outcome, 1 = present and 0 = absent

C1 1st confounder, 1 = present and 0 = absent

C2 2nd confounder, 1 = present and 0 = absent

C3 3rd confounder, 1 = present and 0 = absent

S selection, 1 = selected into the study and 0 = not selected into the study

---

df_uc *Simulated data with uncontrolled confounding*

Description

Data containing one source of bias, three known confounders, and 100,000 observations. This data is obtained from df_uc_source by removing the column *U*. The resulting data corresponds to what a researcher would see in the real-world: information on known confounders (*C1*, *C2*, and *C3*), but not for confounder *U*. As seen in df_uc_source, the true, unbiased exposure-outcome effect estimate = 2.

Usage

```
df_uc
```

Format

A dataframe with 100,000 rows and 7 columns:

X_bi binary exposure, 1 = present and 0 = absent

X_cont continuous exposure

Y_bi binary outcome corresponding to exposure *X_bi*, 1 = present and 0 = absent

Y_cont continuous outcome corresponding to exposure *X_cont*

C1 1st confounder, 1 = present and 0 = absent

C2 2nd confounder, 1 = present and 0 = absent

C3 3rd confounder, 1 = present and 0 = absent

`df_uc_em` *Simulated data with uncontrolled confounding and exposure misclassification*

Description

Data containing two sources of bias, three known confounders, and 100,000 observations. This data is obtained from `df_uc_em_source` by removing the columns *X* and *U*. The resulting data corresponds to what a researcher would see in the real-world: a misclassified exposure, *Xstar*, and missing data on a confounder *U*. As seen in `df_uc_em_source`, the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_uc_em
```

Format

A dataframe with 100,000 rows and 5 columns:

Xstar misclassified exposure, 1 = present and 0 = absent

Y outcome, 1 = present and 0 = absent

C1 1st confounder, 1 = present and 0 = absent

C2 2nd confounder, 1 = present and 0 = absent

C3 3rd confounder, 1 = present and 0 = absent

df_uc_em_sel *Simulated data with uncontrolled confounding, exposure misclassification, and selection bias*

Description

Data containing three sources of bias, three known confounders, and 100,000 observations. This data is obtained by sampling with replacement with probability = *S* from `df_uc_em_sel_source` then removing the columns *X*, *U*, and *S*. The resulting data corresponds to what a researcher would see in the real-world: a misclassified exposure, *Xstar*; missing data on a confounder *U*; and missing data for those not selected into the study (*S*=0). As seen in `df_uc_em_sel_source`, the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_uc_em_sel
```

Format

A dataframe with 100,000 rows and 5 columns:

Xstar misclassified exposure, 1 = present and 0 = absent

Y outcome, 1 = present and 0 = absent

C1 1st confounder, 1 = present and 0 = absent

C2 2nd confounder, 1 = present and 0 = absent

C3 3rd confounder, 1 = present and 0 = absent

df_uc_em_sel_source *Data source for* `df_uc_em_sel`

Description

Data with complete information on the three sources of bias, three known confounders, and 100,000 observations. This data is used to derive `df_uc_em_sel` and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with `df_uc_em_sel`. With this source data, the fitted regression $logit(P(Y = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 C1 + \alpha_3 C2 + \alpha_4 C3 + \alpha_5 U$ shows that the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_uc_em_sel_source
```

Format

A dataframe with 100,000 rows and 8 columns:

X true exposure, 1 = present and 0 = absent

Y outcome, 1 = present and 0 = absent

C1 1st confounder, 1 = present and 0 = absent

C2 2nd confounder, 1 = present and 0 = absent

C3 3rd confounder, 1 = present and 0 = absent

U unmeasured confounder, 1 = present and 0 = absent

Xstar misclassified exposure, 1 = present and 0 = absent

S selection, 1 = selected into the study and 0 = not selected into the study

---

`df_uc_em_source` *Data source for* `df_uc_em`

Description

Data with complete information on the two sources of bias, a known confounder, and 100,000 observations. This data is used to derive `df_uc_em` and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with `df_uc_em`. With this source data, the fitted regression $logit(P(Y = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 C1 + \alpha_3 U$ shows that the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_uc_em_source
```

Format

A dataframe with 100,000 rows and 7 columns:

X true exposure, 1 = present and 0 = absent

Y outcome, 1 = present and 0 = absent

C1 1st confounder, 1 = present and 0 = absent

C2 2nd confounder, 1 = present and 0 = absent

C3 3rd confounder, 1 = present and 0 = absent

U unmeasured confounder, 1 = present and 0 = absent

Xstar misclassified exposure, 1 = present and 0 = absent

df_uc_om — *Simulated data with uncontrolled confounding and outcome misclassification*

Description

Data containing two sources of bias, three known confounders, and 100,000 observations. This data is obtained from `df_uc_om_source` by removing the columns *Y* and *U*. The resulting data corresponds to what a researcher would see in the real-world: a misclassified outcome, *Ystar*, and missing data on the binary confounder *U*. As seen in `df_uc_omc_source`, the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_uc_om
```

Format

A dataframe with 100,000 rows and 5 columns:

X exposure, 1 = present and 0 = absent

Ystar misclassified outcome, 1 = present and 0 = absent

C1 1st confounder, 1 = present and 0 = absent

C2 2nd confounder, 1 = present and 0 = absent

C3 3rd confounder, 1 = present and 0 = absent

df_uc_om_sel — *Simulated data with uncontrolled confounding, outcome misclassification, and selection bias*

Description

Data containing three sources of bias, three known confounders, and 100,000 observations. This data is obtained by sampling with replacement with probability = *S* from `df_uc_om_sel_source` then removing the columns *Y*, *U*, and *S*. The resulting data corresponds to what a researcher would see in the real-world: a misclassified outcome, *Ystar*; missing data on a confounder *U*; and missing data for those not selected into the study (*S*=0). As seen in `df_uc_om_sel_source`, the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_uc_om_sel
```

Format

A dataframe with 100,000 rows and 5 columns:

X exposure, 1 = present and 0 = absent

Ystar misclassified outcome, 1 = present and 0 = absent

C1 1st confounder, 1 = present and 0 = absent

C2 2nd confounder, 1 = present and 0 = absent

C3 3rd confounder, 1 = present and 0 = absent

---

df_uc_om_sel_source *Data source for* df_uc_om_sel

Description

Data with complete information on the three sources of bias, three known confounders, and 100,000 observations. This data is used to derive df_uc_om_sel and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with df_uc_om_sel. With this source data, the fitted regression $logit(P(Y = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 C1 + \alpha_3 C2 + \alpha_4 C3 + \alpha_5 U$ shows that the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_uc_om_sel_source
```

Format

A dataframe with 100,000 rows and 8 columns:

X exposure, 1 = present and 0 = absent

Y true outcome, 1 = present and 0 = absent

C1 1st confounder, 1 = present and 0 = absent

C2 2nd confounder, 1 = present and 0 = absent

C3 3rd confounder, 1 = present and 0 = absent

U unmeasured confounder, 1 = present and 0 = absent

Ystar misclassified outcome, 1 = present and 0 = absent

S selection, 1 = selected into the study and 0 = not selected into the study

df_uc_om_source — *Data source for* df_uc_om

Description

Data with complete information on the two sources of bias, three known confounders, and 100,000 observations. This data is used to derive df_uc_om and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with df_uc_om. With this source data, the fitted regression $logit(P(Y = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 C1 + \alpha_3 U$ shows that the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_uc_om_source
```

Format

A dataframe with 100,000 rows and 7 columns:

X exposure, 1 = present and 0 = absent

Y outcome, 1 = present and 0 = absent

C1 1st confounder, 1 = present and 0 = absent

C2 2nd confounder, 1 = present and 0 = absent

C3 3rd confounder, 1 = present and 0 = absent

U unmeasured confounder, 1 = present and 0 = absent

Ystar misclassified outcome, 1 = present and 0 = absent

df_uc_sel — *Simulated data with uncontrolled confounding and selection bias*

Description

Data containing two sources of bias, three known confounders, and 100,000 observations. This data is obtained by sampling with replacement with probability = *S* from df_uc_sel_source then removing the columns *U* and *S*. The resulting data corresponds to what a researcher would see in the real-world: missing data on confounder *U*; and missing data for those not selected into the study (*S*=0). As seen in df_uc_sel_source, the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_uc_sel
```

Format

A dataframe with 100,000 rows and 5 columns:

X exposure, 1 = present and 0 = absent

Y outcome, 1 = present and 0 = absent

C1 1st confounder, 1 = present and 0 = absent

C2 2nd confounder, 1 = present and 0 = absent

C3 3rd confounder, 1 = present and 0 = absent

`df_uc_sel_source` *Data source for* `df_uc_sel`

Description

Data with complete information on the two sources of bias, a known confounder, and 100,000 observations. This data is used to derive `df_uc_sel` and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with `df_uc_sel`. With this source data, the fitted regression $logit(P(Y = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 C1 + \alpha_3 C2 + \alpha_4 C3 + \alpha_5 U$ shows that the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_uc_sel_source
```

Format

A dataframe with 100,000 rows and 7 columns:

X true exposure, 1 = present and 0 = absent

Y outcome, 1 = present and 0 = absent

C1 1st confounder, 1 = present and 0 = absent

C2 2nd confounder, 1 = present and 0 = absent

C3 3rd confounder, 1 = present and 0 = absent

U unmeasured confounder, 1 = present and 0 = absent

S selection, 1 = selected into the study and 0 = not selected into the study

df_uc_source *Data source for* df_uc

Description

Data with complete information on one source of bias, three known confounders, and 100,000 observations. This data is used to derive df_uc and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with df_uc. With this source data, the fitted regression $logit(P(Y=1)) = \alpha_0 + \alpha_1 X + \alpha_2 C1 + \alpha_3 C2 + \alpha_4 C3 + \alpha_5 U$ shows that the true, unbiased exposure-outcome effect estimate = 2 when:

1. g = logit, Y = *Y_bi*, and X = *X_bi* or
2. g = identity, Y = *Y_cont*, X = *X_cont*.

Usage

```
df_uc_source
```

Format

A dataframe with 100,000 rows and 8 columns:

X_bi binary exposure, 1 = present and 0 = absent

X_cont continuous exposure

Y_bi binary outcome corresponding to exposure *X_bi*, 1 = present and 0 = absent

Y_cont continuous outcome corresponding to exposure *X_cont*

C1 1st confounder, 1 = present and 0 = absent

C2 2nd confounder, 1 = present and 0 = absent

C3 3rd confounder, 1 = present and 0 = absent

U uncontrolled confounder, 1 = present and 0 = absent

multibias_adjust *Simultaneously adjust for multiple biases*

Description

multibias_adjust returns the exposure-outcome odds ratio and confidence interval, adjusted for one or more biases.

Usage

```
multibias_adjust(
  data_observed,
  data_validation = NULL,
  bias_params = NULL,
  bootstrap = FALSE,
  bootstrap_reps = 100,
  level = 0.95
)
```

Arguments

`data_observed`	Object of class `data_observed` corresponding to the data to perform bias analysis on.
`data_validation`	Object of class `data_validation` corresponding to the validation data used to adjust for bias in the observed data. The validation data should have data for the same variables as in `data_observed`, plus data for the missing variables leading to bias.
`bias_params`	Object of class 'bias_params' corresponding to the bias parameters used to adjust for bias in the observed data. There must be parameters corresponding to the bias or biases specified in `data_observed`.
`bootstrap`	Boolean for whether to perform bootstrapping to obtain the estimate and confidence interval.
`bootstrap_reps`	Integer number of bootstrap samples to run in bootstrapping.
`level`	Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

Details

Bias adjustment can be performed by inputting either a validation dataset or the necessary bias parameters. Values for the bias parameters can be applied as fixed values or as single draws from a probability distribution (ex: `rnorm(1, mean = 2, sd = 1)`). The latter has the advantage of allowing the researcher to capture the uncertainty in the bias parameter estimates. To incorporate this uncertainty in the estimate and confidence interval, this function should be run in loop across bootstrap samples of the dataframe for analysis. The estimate and confidence interval would then be obtained from the median and quantiles of the distribution of odds ratio estimates.

Value

A list where the first item is the odds ratio estimate of the effect of the exposure on the outcome and the second item is the confidence interval as the vector: (lower bound, upper bound).

Examples

```
# Adjust for exposure misclassification ------------------------------------
df_observed <- data_observed(
  data = df_em,
```

```
  bias = "em",
  exposure = "Xstar",
  outcome = "Y",
  confounders = "C1"
)

# Using validation data
df_validation <- data_validation(
  data = df_em_source,
  true_exposure = "X",
  true_outcome = "Y",
  confounders = "C1",
  misclassified_exposure = "Xstar"
)

multibias_adjust(
  data_observed = df_observed,
  data_validation = df_validation
)

# Using bias_params
bp <- bias_params(coef_list = list(x = c(-2.10, 1.62, 0.63, 0.35)))

multibias_adjust(
  data_observed = df_observed,
  bias_params = bp
)

# Adjust for three biases ---------------------------------------------------
df_observed <- data_observed(
  data = df_uc_om_sel,
  bias = c("uc", "om", "sel"),
  exposure = "X",
  outcome = "Ystar",
  confounders = c("C1", "C2", "C3")
)

# Using validation data
df_validation <- data_validation(
  data = df_uc_om_sel_source,
  true_exposure = "X",
  true_outcome = "Y",
  confounders = c("C1", "C2", "C3", "U"),
  misclassified_outcome = "Ystar",
  selection = "S"
)

multibias_adjust(
  data_observed = df_observed,
  data_validation = df_validation
)

# Using bias_params
```

```
bp1 <- bias_params(
  coef_list = list(
    u = c(-0.32, 0.59, 0.69),
    y = c(-2.85, 0.71, 1.63, 0.40, -0.85, 0.22),
    s = c(0.00, 0.74, 0.19, 0.02, -0.06, 0.02)
  )
)

multibias_adjust(
  data_observed = df_observed,
  bias_params = bp1
)

bp2 <- bias_params(
  coef_list = list(
    u1y0 = c(-0.20, 0.62, 0.01, -0.08, 0.10, -0.15),
    u0y1 = c(-3.28, 0.63, 1.65, 0.42, -0.85, 0.26),
    u1y1 = c(-2.70, 1.22, 1.64, 0.32, -0.77, 0.09),
    s = c(0.00, 0.74, 0.19, 0.02, -0.06, 0.02)
  )
)

# with bootstrapping
## Not run:
multibias_adjust(
  data_observed = df_observed,
  bias_params = bp2,
  bootstrap = TRUE,
  bootstrap_reps = 1000
)

## End(Not run)
```

Index