Package ‘multibias’

August 21, 2024

Type Package

Title Simultaneous Multi-Bias Adjustment

Version 1.5.2

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 7.2.3

Imports dplyr (>= 1.1.3), magrittr (>= 2.0.3), rlang (>= 1.1.1)

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

URL https://github.com/pcbrendel/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 2024-08-21 10:00:09 UTC

Contents

adjust_emc *Adust for exposure misclassification.*

Description

`adjust_emc` returns the exposure-outcome odds ratio and confidence interval, adjusted for exposure misclassificaiton.

Usage

```
adjust_emc(
  data,
  exposure,
  outcome,
  confounders = NULL,
  x_model_coefs,
  level = 0.95
)
```

Arguments

`data`	Dataframe for analysis.
`exposure`	String name of the exposure variable.
`outcome`	String name of the outcome variable.
`confounders`	String name(s) of the confounder(s). A maximum of three confounders is allowed.
`x_model_coefs`	The regression coefficients corresponding to the model: $logit(P(X = 1)) = \delta_0 + \delta_1 X^* + \delta_2 Y + \delta_{2+j} C_j$, where *X* represents the binary true exposure, *X** is the binary misclassified exposure, *Y* is the outcome, *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders. The number of parameters is therefore 3 + *j*.
`level`	Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

Details

Values for the regression coefficients 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_emc(
  evans,
  exposure = "SMK",
  outcome = "CHD",
  confounders = "HPT",
  x_model_coefs = c(qlogis(0.01), log(6), log(2), log(2))
)
```

`adjust_emc_omc` *Adust for exposure misclassification and outcome misclassification.*

Description

`adjust_emc_omc` returns the exposure-outcome odds ratio and confidence interval, adjusted for exposure misclassification and outcome misclassification. Two different options for the bias parameters are available here: 1) parameters from separate models of *X* and *Y* (`x_model_coefs` and `y_model_coefs`) or 2) parameters from a joint model of *X* and *Y* (`x1y0_model_coefs`, `x0y1_model_coefs`, and `x1y1_model_coefs`).

Usage

```
adjust_emc_omc(
  data,
  exposure,
  outcome,
  confounders = NULL,
  x_model_coefs = NULL,
  y_model_coefs = NULL,
  x1y0_model_coefs = NULL,
  x0y1_model_coefs = NULL,
  x1y1_model_coefs = NULL,
  level = 0.95
)
```

Arguments

`data`	Dataframe for analysis.
`exposure`	String name of the exposure variable.
`outcome`	String name of the outcome variable.
`confounders`	String name(s) of the confounder(s). A maximum of three confounders is allowed.
`x_model_coefs`	The regression coefficients corresponding to the model: $logit(P(X = 1)) = \delta_0 + \delta_1 X^* + \delta_2 Y^* + \delta 2 + jC_j$, where *X* represents the binary true exposure, *X** is the binary misclassified exposure, *Y** is the binary misclassified outcome, *C*

represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders. The number of parameters is therefore 3 + *j*.

`y_model_coefs` The regression coefficients corresponding to the model: $logit(P(Y = 1)) = \beta_0 + \beta_1 X + \beta_2 Y^* + \beta 2 + jC_j$, where *Y* represents the binary true exposure, *X* is the binary exposure, *Y* is the binary misclassified outcome, *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders. The number of parameters is therefore 3 + *j*.

`x1y0_model_coefs`
The regression coefficients corresponding to the model: $log(P(X = 1, Y = 0)/P(X = 0, Y = 0)) = \gamma_{1,0} + \gamma_{1,1} X^* + \gamma_{1,2} Y^* + \gamma_{1,2+j} C_j$, where *X* is the binary true exposure, *Y* is the binary true outcome, *X** is the binary misclassified exposure, *Y** is the binary misclassified outcome, *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders.

`x0y1_model_coefs`
The regression coefficients corresponding to the model: $log(P(X = 0, U = 1)/P(X = 0, U = 0)) = \gamma_{2,0} + \gamma_{2,1} X^* + \gamma_{2,2} Y^* + \gamma_{2,2+j} C_j$, where *X* is the binary true exposure, *Y* is the binary true outcome, *X** is the binary misclassified exposure, *Y** is the binary misclassified outcome, *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders.

`x1y1_model_coefs`
The regression coefficients corresponding to the model: $log(P(X = 1, Y = 1)/P(X = 0, Y = 0)) = \gamma_{3,0} + \gamma_{3,1} X^* + \gamma_{3,2} Y^* + \gamma_{3,2+j} C_j$, where *X* is the binary true exposure, *Y* is the binary true outcome, *X** is the binary misclassified exposure, *Y** is the binary misclassified outcome, *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders.

`level` Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

Details

Values for the regression coefficients 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

```
# Using x_model_coefs and y_model_coefs -------------------------------------
```

```
adjust_emc_omc(
  df_emc_omc,
  exposure = "Xstar",
  outcome = "Ystar",
  confounders = "C1",
  x_model_coefs = c(-2.15, 1.64, 0.35, 0.38),
  y_model_coefs = c(-3.10, 0.63, 1.60, 0.39)
)

# Using x1y0_model_coefs, x0y1_model_coefs, and x1y1_model_coefs ------------
adjust_emc_omc(
  df_emc_omc,
  exposure = "Xstar",
  outcome = "Ystar",
  confounders = "C1",
  x1y0_model_coefs = c(-2.18, 1.63, 0.23, 0.36),
  x0y1_model_coefs = c(-3.17, 0.22, 1.60, 0.40),
  x1y1_model_coefs = c(-4.76, 1.82, 1.83, 0.72)
)
```

`adjust_emc_sel` *Adust for exposure misclassification and selection bias.*

Description

`adjust_emc_sel` returns the exposure-outcome odds ratio and confidence interval, adjusted for exposure misclassification and selection bias.

Usage

```
adjust_emc_sel(
  data,
  exposure,
  outcome,
  confounders = NULL,
  x_model_coefs,
  s_model_coefs,
  level = 0.95
)
```

Arguments

`data`	Dataframe for analysis.
`exposure`	String name of the exposure variable.
`outcome`	String name of the outcome variable.
`confounders`	String name(s) of the confounder(s). A maximum of three confounders is allowed.

`x_model_coefs` The regression coefficients corresponding to the model: $logit(P(X = 1)) = \delta_0 + \delta_1 X^* + \delta_2 Y + \delta 2 + jC_j$, where *X* represents the binary true exposure, *X** is the binary misclassified exposure, *Y* is the outcome, *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders. The number of parameters is therefore 3 + *j*.

`s_model_coefs` The regression coefficients corresponding to the model: $logit(P(S = 1)) = \beta_0 + \beta_1 X^* + \beta_2 Y + \beta 2 + jC_j$, where *S* represents binary selection, *X** is the binary misclassified exposure, *Y* is the outcome, *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders. The number of parameters is therefore 3 + *j*.

`level` Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

Details

Values for the regression coefficients 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_emc_sel(
  df_emc_sel,
  exposure = "Xstar",
  outcome = "Y",
  confounders = "C1",
  x_model_coefs = c(-2.78, 1.62, 0.58, 0.34),
  s_model_coefs = c(0.04, 0.18, 0.92, 0.05)
)
```

`adjust_multinom_uc_emc_sel`

Adust for uncontrolled confounding, exposure misclassification, and selection bias.

Description

`adjust_multinom_uc_emc_sel` returns the exposure-outcome odds ratio and confidence interval, adjusted for uncontrolled confounding, exposure misclassificaiton, and selection bias.

Usage

```
adjust_multinom_uc_emc_sel(
  data,
  exposure,
  outcome,
  confounders = NULL,
  x1u0_model_coefs,
  x0u1_model_coefs,
  x1u1_model_coefs,
  s_model_coefs,
  level = 0.95
)
```

Arguments

`data` Dataframe for analysis.

`exposure` String name of the exposure variable.

`outcome` String name of the outcome variable.

`confounders` String name(s) of the confounder(s). A maximum of three confounders is allowed.

`x1u0_model_coefs` The regression coefficients corresponding to the model: $log(P(X = 1, U = 0)/P(X = 0, U = 0)) = \gamma_{1,0} + \gamma_{1,1}X^* + \gamma_{1,2}Y + \gamma_{1,2+j}C_j$, where *X* is the binary true exposure, *U* is the binary unmeasured confounder, *X** is the binary misclassified exposure, *Y* is the binary outcome, *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders.

`x0u1_model_coefs` The regression coefficients corresponding to the model: $log(P(X = 0, U = 1)/P(X = 0, U = 0)) = \gamma_{2,0} + \gamma_{2,1}X^* + \gamma_{2,2}Y + \gamma_{2,2+j}C_j$, where *X* is the binary true exposure, *U* is the binary unmeasured confounder, *X** is the binary misclassified exposure, *Y* is the binary outcome, *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders.

`x1u1_model_coefs` The regression coefficients corresponding to the model: $log(P(X = 1, U = 1)/P(X = 0, U = 0)) = \gamma_{3,0} + \gamma_{3,1}X^* + \gamma_{3,2}Y + \gamma_{3,2+j}C_j$, where *X* is the binary true exposure, *U* is the binary unmeasured confounder, *X** is the binary misclassified exposure, *Y* is the binary outcome, *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders.

`s_model_coefs` The regression coefficients corresponding to the model: $logit(P(S = 1)) = \beta_0 + \beta_1 X^* + \beta_2 Y + \beta_{2+j}C_j$, where *S* represents binary selection, *X** is the binary misclassified exposure, *Y* is the binary outcome, *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders.

`level` Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

Details

This function uses one bias model, a multinomial logistic regression model, to predict the uncontrolled confounder (U) and exposure (X). If separate bias models for X and U are desired, use `adjust_uc_emc_sel`.

Values for the regression coefficients 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_multinom_uc_emc_sel(
  df_uc_emc_sel,
  exposure = "Xstar",
  outcome = "Y",
  confounders = c("C1", "C2", "C3"),
  x1u0_model_coefs = c(-2.78, 1.62, 0.61, 0.36, -0.27, 0.88),
  x0u1_model_coefs = c(-0.17, -0.01, 0.71, -0.08, 0.07, -0.15),
  x1u1_model_coefs = c(-2.36, 1.62, 1.29, 0.25, -0.06, 0.74),
  s_model_coefs = c(0.00, 0.26, 0.78, 0.03, -0.02, 0.10)
)
```

---

adjust_multinom_uc_omc_sel

Adust for uncontrolled confounding, outcome misclassification, and selection bias.

Description

`adjust_multinom_uc_omc_sel` returns the exposure-outcome odds ratio and confidence interval, adjusted for uncontrolled confounding, outcome misclassificaiton, and selection bias.

Usage

```
adjust_multinom_uc_omc_sel(
  data,
  exposure,
  outcome,
  confounders = NULL,
```

```
  u0y1_model_coefs,
  u1y0_model_coefs,
  u1y1_model_coefs,
  s_model_coefs,
  level = 0.95
)
```

Arguments

`data` Dataframe for analysis.

`exposure` String name of the exposure variable.

`outcome` String name of the outcome variable.

`confounders` String name(s) of the confounder(s). A maximum of three confounders is allowed.

`u0y1_model_coefs` The regression coefficients corresponding to the model: $log(P(U = 0, Y = 1)/P(U = 0, Y = 0)) = \gamma_{2,0} + \gamma_{2,1}X + \gamma_{2,2}Y^* + \gamma_{2,2+j}C_j$, where U is the binary unmeasured confounder, Y is the binary true outcome, X is the binary exposure, Y* is the binary misclassified outcome, C represents the vector of binary measured confounders (if any), and j corresponds to the number of measured confounders.

`u1y0_model_coefs` The regression coefficients corresponding to the model: $log(P(U = 1, Y = 0)/P(U = 0, Y = 0)) = \gamma_{1,0} + \gamma_{1,1}X + \gamma_{1,2}Y^* + \gamma_{1,2+j}C_j$, where *U* is the binary unmeasured confounder, *Y* is the binary true outcome, *X* is the binary exposure, *Y** is the binary misclassified outcome, *C* represents the vector of binary measured confounders (if any), and *j* corresponds to the number of measured confounders.

`u1y1_model_coefs` The regression coefficients corresponding to the model: $log(P(U = 1, Y = 1)/P(U = 0, Y = 0)) = \gamma_{3,0} + \gamma_{3,1}X + \gamma_{3,2}Y^* + \gamma_{3,2+j}C_j$, where *U* is the binary unmeasured confounder, *Y* is the binary true outcome, *X* is the binary exposure, *Y** is the binary misclassified outcome, *C* represents the vector of binary measured confounders (if any), and *j* corresponds to the number of measured confounders.

`s_model_coefs` The regression coefficients corresponding to the model: $logit(P(S = 1)) = \beta_0 + \beta_1 X + \beta_2 Y^* + \beta_{2+j}C_j$, where *S* represents binary selection, *X* is the binary exposure, *Y** is the binary misclassified outcome, *C* represents the vector of binary measured confounders (if any), and *j* corresponds to the number of measured confounders.

`level` Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

Details

This function uses one bias model, a multinomial logistic regression model, to predict the uncontrolled confounder (*U*) and outcome (*Y*). If separate bias models for *U* and *Y* are desired, use `adjust_uc_omc_sel`.

Values for the regression coefficients 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_multinom_uc_omc_sel(
  df_uc_omc_sel,
  exposure = "X",
  outcome = "Ystar",
  confounders = c("C1", "C2", "C3"),
  u1y0_model_coefs = c(-0.20, 0.62, 0.01, -0.08, 0.10, -0.15),
  u0y1_model_coefs = c(-3.28, 0.63, 1.65, 0.42, -0.85, 0.26),
  u1y1_model_coefs = c(-2.70, 1.22, 1.64, 0.32, -0.77, 0.09),
  s_model_coefs = c(0.00, 0.74, 0.19, 0.02, -0.06, 0.02)
)
```

`adjust_omc` *Adust for outcome misclassification.*

Description

`adjust_omc` returns the exposure-outcome odds ratio and confidence interval, adjusted for outcome misclassificaiton.

Usage

```
adjust_omc(
  data,
  exposure,
  outcome,
  confounders = NULL,
  y_model_coefs,
  level = 0.95
)
```

Arguments

`data` Dataframe for analysis.

`exposure` String name of the exposure variable.

`outcome` String name of the outcome variable.

`confounders` String name(s) of the confounder(s). A maximum of three confounders is allowed.

`y_model_coefs` The regression coefficients corresponding to the model: $logit(P(Y = 1)) = \_delta_0 + \_delta_1 X + \_delta_2 Y^* + \_delta_{2+j} C_j$, where *Y* represents the binary true outcome, *X* is the exposure, *Y** is the binary misclassified outcome, *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders. The number of parameters is therefore 3 + *j*.

`level` Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

Details

Values for the regression coefficients 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_omc(
  evans,
  exposure = "SMK",
  outcome = "CHD",
  confounders = "HPT",
  y_model_coefs = c(qlogis(0.01), log(1.5), log(5), log(1.5))
)
```

`adjust_omc_sel` *Adust for outcome misclassification and selection bias.*

Description

`adjust_omc_sel` returns the exposure-outcome odds ratio and confidence interval, adjusted for outcome misclassification and selection bias.

Usage

```
adjust_omc_sel(
  data,
  exposure,
  outcome,
  confounders = NULL,
  y_model_coefs,
  s_model_coefs,
  level = 0.95
)
```

Arguments

`data`	Dataframe for analysis.
`exposure`	String name of the exposure variable.
`outcome`	String name of the outcome variable.
`confounders`	String name(s) of the confounder(s). A maximum of three confounders is allowed.
`y_model_coefs`	The regression coefficients corresponding to the model: $logit(P(Y = 1)) = \delta_0 + \delta_1 X + \delta_2 Y^* + \delta_{2+j} C_j$, where *Y* represents the binary true outcome, *X* is the exposure, *Y** is the binary misclassified outcome, *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders. The number of parameters is therefore 3 + *j*.
`s_model_coefs`	The regression coefficients corresponding to the model: $logit(P(S = 1)) = \beta_0 + \beta_1 X + \beta_2 Y^* + \beta_{2+j} C_j$, where *S* represents binary selection, *X* is the exposure, *Y** is the binary misclassified outcome, *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders. The number of parameters is therefore 3 + *j*.
`level`	Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

Details

Values for the regression coefficients 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_omc_sel(
  df_omc_sel,
  exposure = "X",
  outcome = "Ystar",
  confounders = "C1",
  y_model_coefs = c(-3.24, 0.58, 1.59, 0.45),
  s_model_coefs = c(0.03, 0.92, 0.12, 0.05)
)
```

---

adjust_sel *Adust for selection bias.*

Description

`adjust_sel` returns the exposure-outcome odds ratio and confidence interval, adjusted for selection bias.

Usage

```
adjust_sel(
  data,
  exposure,
  outcome,
  confounders = NULL,
  s_model_coefs,
  level = 0.95
)
```

Arguments

data	Dataframe for analysis.
exposure	String name of the exposure variable.
outcome	String name of the outcome variable.
confounders	String name(s) of the confounder(s). A maximum of three confounders is allowed.
s_model_coefs	The regression coefficients corresponding to the model: $logit(P(S = 1)) = \beta_0 + \beta_1 X + \beta_2 Y$, where *S* represents binary selection, *X* is the exposure, and *Y* is the outcome. The number of parameters is therefore 3.
level	Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

Details

Values for the regression coefficients 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_sel(
  evans,
  exposure = "SMK",
  outcome = "CHD",
  confounders = "HPT",
  s_model_coefs = c(qlogis(0.25), log(0.75), log(0.75))
)
```

`adjust_uc` *Adust for uncontrolled confounding.*

Description

`adjust_uc` returns the exposure-outcome odds ratio and confidence interval, adjusted for uncontrolled confounding from a binary confounder.

Usage

```
adjust_uc(
  data,
  exposure,
  outcome,
  confounders = NULL,
  u_model_coefs,
  level = 0.95
)
```

Arguments

`data`	Dataframe for analysis.
`exposure`	String name of the exposure variable.
`outcome`	String name of the outcome variable.
`confounders`	String name(s) of the confounder(s). A maximum of three confounders is allowed.
`u_model_coefs`	The regression coefficients corresponding to the model: $logit(P(U = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 Y + \alpha_{2+j} C_j$, where *U* is the binary unmeasured confounder, *X* is the exposure, *Y* is the outcome, *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders. The number of parameters therefore equals 3 + *j*.
`level`	Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

Details

Values for the regression coefficients 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_uc(
  evans,
  exposure = "SMK",
  outcome = "CHD",
  confounders = "HPT",
  u_model_coefs = c(qlogis(0.25), log(0.5), log(2.5), log(2)),
)
```

`adjust_uc_emc` *Adust for uncontrolled confounding and exposure misclassification.*

Description

`adjust_uc_emc` returns the exposure-outcome odds ratio and confidence interval, adjusted for uncontrolled confounding and exposure misclassificaiton. Two different options for the bias parameters are available here: 1) parameters from separate models of *U* and *X* (`u_model_coefs` and `x_model_coefs`) or 2) parameters from a joint model of *U* and *X* (`x1u0_model_coefs`, `x0u1_model_coefs`, and `x1u1_model_coefs`).

Usage

```
adjust_uc_emc(
  data,
  exposure,
  outcome,
  confounders = NULL,
  u_model_coefs = NULL,
  x_model_coefs = NULL,
  x1u0_model_coefs = NULL,
  x0u1_model_coefs = NULL,
  x1u1_model_coefs = NULL,
  level = 0.95
)
```

Arguments

`data` Dataframe for analysis.

`exposure` String name of the exposure variable.

`outcome` String name of the outcome variable.

`confounders` String name(s) of the confounder(s). A maximum of three confounders is allowed.

`u_model_coefs` The regression coefficients corresponding to the model: $logit(P(U = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 Y$, where *U* is the binary unmeasured confounder, *X* is the binary true exposure, and *Y* is the outcome. The number of parameters therefore equals 3.

`x_model_coefs` The regression coefficients corresponding to the model: $logit(P(X = 1)) = \delta_0 + \delta_1 X^* + \delta_2 Y + \delta_{2+j} C_j$, where *X* represents the binary true exposure, *X** is the binary misclassified exposure, *Y* is the outcome, and *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders. The number of parameters therefore equals 3 + *j*.

`x1u0_model_coefs`
The regression coefficients corresponding to the model: $log(P(X = 1, U = 0)/P(X = 0, U = 0)) = \gamma_{1,0} + \gamma_{1,1} X^* + \gamma_{1,2} Y + \gamma_{1,2+j} C_j$, where *X* is the binary true exposure, *U* is the binary unmeasured confounder, *X** is the binary misclassified exposure, *Y* is the outcome, *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders.

`x0u1_model_coefs`
The regression coefficients corresponding to the model: $log(P(X = 0, U = 1)/P(X = 0, U = 0)) = \gamma_{2,0} + \gamma_{2,1} X^* + \gamma_{2,2} Y + \gamma_{2,2+j} C_j$, where *X* is the

binary true exposure, *U* is the binary unmeasured confounder, *X** is the binary misclassified exposure, *Y* is the outcome, *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders.

`x1u1_model_coefs`
: The regression coefficients corresponding to the model: $log(P(X = 1, U = 1)/P(X = 0, U = 0)) = \gamma_{3,0} + \gamma_{3,1}X^* + \gamma_{3,2}Y + \gamma_{3,2+j}C_j$, where *X* is the binary true exposure, *U* is the binary unmeasured confounder, *X** is the binary misclassified exposure, *Y* is the outcome, *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders.

`level`
: Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

Details

Values for the regression coefficients 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

```
# Using u_model_coefs and x_model_coefs -------------------------------------
adjust_uc_emc(
  df_uc_emc,
  exposure = "Xstar",
  outcome = "Y",
  confounders = "C1",
  u_model_coefs = c(-0.23, 0.63, 0.66),
  x_model_coefs = c(-2.47, 1.62, 0.73, 0.32)
)

# Using x1u0_model_coefs, x0u1_model_coefs, x1u1_model_coefs ----------------
adjust_uc_emc(
  df_uc_emc,
  exposure = "Xstar",
  outcome = "Y",
  confounders = "C1",
  x1u0_model_coefs = c(-2.82, 1.62, 0.68, -0.06),
  x0u1_model_coefs = c(-0.20, 0.00, 0.68, -0.05),
  x1u1_model_coefs = c(-2.36, 1.62, 1.29, 0.27)
)
```

adjust_uc_emc_sel	*Adust for uncontrolled confounding, exposure misclassification, and selection bias.*

Description

`adjust_uc_emc_sel` returns the exposure-outcome odds ratio and confidence interval, adjusted for uncontrolled confounding, exposure misclassificaiton, and selection bias.

Usage

```
adjust_uc_emc_sel(
  data,
  exposure,
  outcome,
  confounders = NULL,
  u_model_coefs,
  x_model_coefs,
  s_model_coefs,
  level = 0.95
)
```

Arguments

`data`	Dataframe for analysis.
`exposure`	String name of the exposure variable.
`outcome`	String name of the outcome variable.
`confounders`	String name(s) of the confounder(s). A maximum of three confounders is allowed.
`u_model_coefs`	The regression coefficients corresponding to the model: $logit(P(U = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 Y$, where *U* is the binary unmeasured confounder, *X* is the binary true exposure, and *Y* is the binary outcome. The number of parameters therefore equals 3.
`x_model_coefs`	The regression coefficients corresponding to the model: $logit(P(X = 1)) = \delta_0 + \delta_1 X^* + \delta_2 Y + \delta_{2+j} C_j$, where *X* represents binary true exposure, *X** is the binary misclassified exposure, *Y* is the binary outcome, *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders. The number of parameters therefore equals 3 + *j*.
`s_model_coefs`	The regression coefficients corresponding to the model: $logit(P(S = 1)) = \beta_0 + \beta_1 X^* + \beta_2 Y + \beta_{2+j} C_j$, where *S* represents binary selection, *X** is the binary misclassified exposure, *Y* is the binary outcome, *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders. The number of parameters therefore equals 3 + *j*.
`level`	Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

Details

This function uses two separate logistic regression models to predict the uncontrolled confounder (U) and exposure (X). If a single bias model for jointly modeling X and U is desired use `adjust_multinom_uc_emc_sel`.

Values for the regression coefficients 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_uc_emc_sel(
  df_uc_emc_sel,
  exposure = "Xstar",
  outcome = "Y",
  confounders = c("C1", "C2", "C3"),
  u_model_coefs = c(-0.32, 0.59, 0.69),
  x_model_coefs = c(-2.44, 1.62, 0.72, 0.32, -0.15, 0.85),
  s_model_coefs = c(0.00, 0.26, 0.78, 0.03, -0.02, 0.10)
)
```

---

adjust_uc_omc — *Adust for uncontrolled confounding and outcome misclassification.*

Description

`adjust_uc_omc` returns the exposure-outcome odds ratio and confidence interval, adjusted for uncontrolled confounding and outcome misclassificaiton. Two different options for the bias parameters are available here: 1) parameters from separate models of U and Y (`u_model_coefs` and `y_model_coefs`) or 2) parameters from a joint model of U and Y (`u1y0_model_coefs`, `u0y1_model_coefs`, and `u1y1_model_coefs`).

Usage

```
adjust_uc_omc(
  data,
  exposure,
  outcome,
  confounders = NULL,
  u_model_coefs = NULL,
```

```
  y_model_coefs = NULL,
  u1y0_model_coefs = NULL,
  u0y1_model_coefs = NULL,
  u1y1_model_coefs = NULL,
  level = 0.95
)
```

Arguments

`data` Dataframe for analysis.

`exposure` String name of the exposure variable.

`outcome` String name of the outcome variable.

`confounders` String name(s) of the confounder(s). A maximum of three confounders is allowed.

`u_model_coefs` The regression coefficients corresponding to the model: $logit(P(U = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 Y$, where *U* is the binary unmeasured confounder, *X* is the exposure, *Y* is the binary true outcome. The number of parameters therefore equals 3.

`y_model_coefs` The regression coefficients corresponding to the model: $logit(P(Y = 1)) = \delta_0 + \delta_1 X + \delta_2 Y^* + \delta_{2+j} C_j$, where *Y* represents binary true outcome, *X* is the exposure, *Y** is the binary misclassified outcome, *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders. The number of parameters therefore equals 3 + *j*.

`u1y0_model_coefs`
The regression coefficients corresponding to the model: $log(P(U = 1, Y = 0)/P(U = 0, Y = 0)) = \gamma_{1,0} + \gamma_{1,1} X + \gamma_{1,2} Y^* + \gamma_{1,2+j} C_j$, where *U* is the binary unmeasured confounder, *Y* is the binary true outcome, *X* is the exposure, *Y** is the binary misclassified outcome, *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders.

`u0y1_model_coefs`
The regression coefficients corresponding to the model: $log(P(U = 0, Y = 1)/P(U = 0, Y = 0)) = \gamma_{2,0} + \gamma_{2,1} X + \gamma_{2,2} Y^* + \gamma_{2,2+j} C_j$, where *U* is the binary unmeasured confounder, *Y* is the binary true outcome, *X* is the exposure, *Y** is the binary misclassified outcome, *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders.

`u1y1_model_coefs`
The regression coefficients corresponding to the model: $log(P(U = 1, Y = 1)/P(U = 0, Y = 0)) = \gamma_{3,0} + \gamma_{3,1} X + \gamma_{3,2} Y^* + \gamma_{3,2+j} C_j$, where *U* is the binary unmeasured confounder, *Y* is the binary true outcome, *X* is the exposure, *Y** is the binary misclassified outcome, *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders.

`level` Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

Details

Values for the regression coefficients 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

```
# Using u_model_coefs and y_model_coefs -------------------------------------
adjust_uc_omc(
  df_uc_omc,
  exposure = "X",
  outcome = "Ystar",
  confounders = "C1",
  u_model_coefs = c(-0.22, 0.61, 0.70),
  y_model_coefs = c(-2.85, 0.73, 1.60, 0.38)
)

# Using u1y0_model_coefs, u0y1_model_coefs, u1y1_model_coefs ----------------
adjust_uc_omc(
  df_uc_omc,
  exposure = "X",
  outcome = "Ystar",
  confounders = "C1",
  u1y0_model_coefs = c(-0.19, 0.61, 0.00, -0.07),
  u0y1_model_coefs = c(-3.21, 0.60, 1.60, 0.36),
  u1y1_model_coefs = c(-2.72, 1.24, 1.59, 0.34)
)
```

`adjust_uc_omc_sel` *Adust for uncontrolled confounding, outcome misclassification, and selection bias.*

Description

`adjust_uc_omc_sel` returns the exposure-outcome odds ratio and confidence interval, adjusted for uncontrolled confounding, outcome misclassificaiton, and selection bias.

Usage

```
adjust_uc_omc_sel(
  data,
  exposure,
  outcome,
```

```
  confounders = NULL,
  u_model_coefs,
  y_model_coefs,
  s_model_coefs,
  level = 0.95
)
```

Arguments

`data`	Dataframe for analysis.
`exposure`	String name of the exposure variable.
`outcome`	String name of the outcome variable.
`confounders`	String name(s) of the confounder(s). A maximum of three confounders is allowed.
`u_model_coefs`	The regression coefficients corresponding to the model: $logit(P(U = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 Y$, where *U* is the binary unmeasured confounder, *X* is the binary exposure, and *Y* is the binary true outcome. The number of parameters therefore equals 3.
`y_model_coefs`	The regression coefficients corresponding to the model: $logit(P(Y = 1)) = \delta_0 + \delta_1 X + \delta_2 Y^* + \delta_{2+j} C_j$, where *Y* represents binary true outcome, *X* is the binary exposure, *Y** is the binary misclassified outcome, *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders. The number of parameters therefore equals 3 + *j*.
`s_model_coefs`	The regression coefficients corresponding to the model: $logit(P(S = 1)) = \beta_0 + \beta_1 X + \beta_2 Y^* + \beta_{2+j} C_j$, where *S* represents binary selection, *X* is the binary exposure, *Y** is the binary misclassified outcome, *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders. The number of parameters therefore equals 3 + *j*.
`level`	Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

Details

This function uses two separate logistic regression models to predict the uncontrolled confounder (U) and outcome (Y). If a single bias model for jointly modeling *Y* and *U* is desired use `adjust_multinom_uc_omc_sel`.

Values for the regression coefficients 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_uc_omc_sel(
  df_uc_omc_sel,
  exposure = "X",
  outcome = "Ystar",
  confounders = c("C1", "C2", "C3"),
  u_model_coefs = c(-0.32, 0.59, 0.69),
  y_model_coefs = c(-2.85, 0.71, 1.63, 0.40, -0.85, 0.22),
  s_model_coefs = c(0.00, 0.74, 0.19, 0.02, -0.06, 0.02)
)
```

`adjust_uc_sel` *Adust for uncontrolled confounding and selection bias.*

Description

`adjust_uc_sel` returns the exposure-outcome odds ratio and confidence interval, adjusted for uncontrolled confounding and exposure misclassificaiton.

Usage

```
adjust_uc_sel(
  data,
  exposure,
  outcome,
  confounders = NULL,
  u_model_coefs,
  s_model_coefs,
  level = 0.95
)
```

Arguments

`data`	Dataframe for analysis.
`exposure`	String name of the exposure variable.
`outcome`	String name of the outcome variable.
`confounders`	String name(s) of the confounder(s). A maximum of three confounders is allowed.
`u_model_coefs`	The regression coefficients corresponding to the model: $logit(P(U = 1)) = \alpha_0 + \alpha_1 X + \alpha_2 Y + \alpha_{2+j} C_j$, where *U* is the binary unmeasured confounder, *X* is the exposure, *Y* is the outcome, *C* represents the vector of measured confounders (if any), and *j* corresponds to the number of measured confounders. The number of parameters therefore equals 3 + *j*.
`s_model_coefs`	The regression coefficients corresponding to the model: $logit(P(S = 1)) = \beta_0 + \beta_1 X + \beta_2 Y$, where *S* represents binary selection, *X* is the exposure, and *Y* is the outcome. The number of parameters therefore equals 3.

`level` Value from 0-1 representing the full range of the confidence interval. Default is 0.95.

Details

Values for the regression coefficients 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_uc_sel(
  df_uc_sel,
  exposure = "X",
  outcome = "Y",
  confounders = c("C1", "C2", "C3"),
  u_model_coefs = c(-0.19, 0.61, 0.72, -0.09, 0.10, -0.15),
  s_model_coefs = c(-0.01, 0.92, 0.94)
)
```

`df_emc` *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_emc
```

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_emc_omc` *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_emc_omc_source`, the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_emc_omc
```

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_emc_omc_source` *Data source for* `df_emc_omc`

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_emc_omc` and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with `df_emc_omc`. 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_emc_omc_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_emc_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_emc_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_emc_sel_source`, the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_emc_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_emc_sel_source *Data source for* df_emc_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_emc_sel` and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with `df_emc_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_emc_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_emc_source *Data source for* df_emc

Description

Data with complete information on one sources of bias, three known confounders, and 100,000 observations. This data is used to derive `df_emc` and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with `df_emc`. 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_emc_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_omc` *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_omc_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_omc_source`, the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_omc
```

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_omc_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_omc_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_omc_sel_source`, the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_omc_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_omc_sel_source *Data source for* `df_omc_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_omc_sel` and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with `df_omc_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_omc_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_omc_source *Data source for* `df_omc`

Description

Data with complete information on one sources of bias, three known confounders, and 100,000 observations. This data is used to derive `df_omc` and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with `df_omc`. 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_omc_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_emc — *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_emc_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_emc_source`, the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_uc_emc
```

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_emc_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_emc_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_emc_sel_source`, the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_uc_emc_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_emc_sel_source` *Data source for* `df_uc_emc_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_emc_sel` and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with `df_uc_emc_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_emc_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_emc_source — *Data source for* df_uc_emc

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_emc and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with df_uc_emc. 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_emc_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_omc — *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_omc_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_omc
```

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_omc_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_omc_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_omc_sel_source`, the true, unbiased exposure-outcome odds ratio = 2.

Usage

```
df_uc_omc_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_omc_sel_source *Data source for* df_uc_omc_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_omc_sel and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with df_uc_omc_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_omc_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_omc_source *Data source for* df_uc_omc

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_omc and can be used to obtain bias parameters for purposes of validating the simultaneous multi-bias adjustment method with df_uc_omc. 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_omc_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

`evans` *Evans County dataset*

Description

Data from a cohort study in which white males in Evans County were followed for 7 years, with coronary heart disease as the outcome of interest.

Usage

```
evans
```

Format

A dataframe with 609 rows and 9 columns:

ID subject identifiction

CHD outcome variable; 1 = coronary heart disease

AGE age (in years)

CHL cholesterol, mg/dl

SMK 1 = subject has ever smoked

ECG 1 = presence of electrocardiogram abnormality

DBP diastolic blood pressure, mmHg

SBP systolic blood pressure, mmHg

HPT 1 = SBP greater than or equal to 160 or DBP greater than or equal to 95

Source

<http://web1.sph.emory.edu/dkleinb/logreg3.htm#data>

Index