# Package ‘ExactMed’

January 20, 2025

**Type** Package

**Title** Exact Mediation Analysis for Binary Outcomes

**Version** 0.3.0

**Description** A tool for conducting exact parametric regression-based causal mediation analysis of binary outcomes as described in Samoilenko, Blais and Lefebvre (2018) <doi:10.1353/obs.2018.0013>; Samoilenko, Lefebvre (2021) <doi:10.1093/aje/kwab055>; and Samoilenko, Lefebvre (2023) <doi:10.1002/sim.9621>.

**Depends** R (>= 3.5.0)

**Imports** brglm2, stats, utils, sandwich, lmtest, pkgcond, mlogit, dfidx, nnet

**License** GPL-3

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.2.3

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

**VignetteBuilder** knitr

**Config/testthat/edition** 3

**URL** `https://caubm.github.io/ExactMed/`

**BugReports** `https://github.com/caubm/ExactMed/issues`

**NeedsCompilation** no

**Author** Miguel Caubet [aut, cre],
Mariia Samoilenko [aut],
Jesse Gervais [aut],
Geneviève Lefebvre [aut]

**Maintainer** Miguel Caubet <`miguelcaubet@gmail.com`>

**Repository** CRAN

**Date/Publication** 2023-09-22 15:10:06 UTC

# Contents

---

## `datamed` *Data for Examples (Binary Mediator)*

### Description

Simulated data set containing 1000 observations on 5 measured variables with no missing values. The first three variables are the binary exposure, mediator and outcome, respectively, while the last two variables are the potential adjustment covariates (one binary and one continuous).

### Usage

```
data(datamed)
```

### Format

A data frame with 1000 rows and 5 variables:

**X** exposure, binary variable

**M** mediator, binary variable

**Y** outcome, binary variable

**C1** first covariate, binary variable

**C2** second covariate, continuous variable

---

## `datamed_c` *Data for Examples (Continuous Mediator)*

### Description

Simulated data set containing 1000 observations on 5 measured variables with no missing values. The first three variables are the binary exposure, the continuous mediator and the binary outcome, respectively, while the last two variables are the potential adjustment covariates (one binary and one continuous).

### Usage

`data(datamed_c)`

### Format

A data frame with 1000 rows and 5 variables:

**X** exposure, binary variable

**M** mediator, continuous variable

**Y** outcome, binary variable

**C1** first covariate, binary variable

**C2** second covariate, continuous variable

---

## `datamed_cat` *Data for Examples (Categorical Mediator)*

### Description

Simulated data set containing 1000 observations on 5 measured variables with no missing values. The first three variables are the binary exposure, the categorical mediator and the binary outcome, respectively, while the last two variables are the potential adjustment covariates (one binary and one continuous).

### Usage

`data(datamed_cat)`

### Format

A data frame with 1000 rows and 5 variables:

**X** exposure, binary variable

**M** mediator, categorical variable

**Y** outcome, binary variable

**C1** first covariate, binary variable

**C2** second covariate, continuous variable

---

## `exactmed` *Exact Mediation Effects Computation (Binary Mediator)*

---

### Description

Relying on a regression-based approach, the `exactmed()` function calculates standard causal mediation effects when the outcome and the mediator are binary. More precisely, `exactmed()` uses a logistic regression specification for both the outcome and the mediator in order to compute *exact* conditional natural direct and indirect effects (see details in Samoilenko and Lefebvre, 2021). The function returns point and interval estimates for the conditional natural effects without making any assumption regarding the rareness or commonness of the outcome (hence the term exact). For completeness, `exactmed()` also calculates the conditional controlled direct effects at both values of the mediator. Natural and controlled effects estimates are reported using three different scales: odds ratio (OR), risk ratio (RR) and risk difference (RD). The interval estimates can be obtained either by the delta method or the bootstrap.

### Usage

```
exactmed(
  data,
  a,
  m,
  y,
  a1,
  a0,
  m_cov = NULL,
  y_cov = NULL,
  m_cov_cond = NULL,
  y_cov_cond = NULL,
  adjusted = TRUE,
  interaction = TRUE,
  Firth = FALSE,
  boot = FALSE,
  nboot = 1000,
  bootseed = 1991,
  confcoef = 0.95,
  hvalue_m = NULL,
  hvalue_y = NULL,
  yprevalence = NULL
)
```

### Arguments

| | |
|---|---|
| `data` | A named data frame that includes the exposure, mediator and outcome variables as well as the covariates to be adjusted for in the models. The exposure can be either binary or continuous. If a covariate is categorical, it has to be included in the data frame as a factor, character or logical variable. |

| | |
|---|---|
| `a` | The name of the binary or continuous exposure variable. |
| `m` | The name of the binary mediator variable. |
| `y` | The name of the binary outcome variable. |
| `a1` | A value corresponding to the high level of the exposure. |
| `a0` | A value corresponding to the low level of the exposure. |
| `m_cov` | A vector containing the names of the adjustment variables (covariates) in the mediator model. |
| `y_cov` | A vector containing the names of the adjustment variables (covariates) in the outcome model. |
| `m_cov_cond` | A named vector (atomic vector or list) containing specific values for some or all of the adjustment covariates `m_cov` in the mediator model. Please consult the package vignette for details. |
| `y_cov_cond` | A named vector (atomic vector or list) containing specific values for some or all of the adjustment covariates `y_cov` in the outcome model. Please consult the package vignette for details. |
| `adjusted` | A logical variable specifying whether to obtain adjusted or unadjusted estimates. If `adjusted = FALSE`, vectors `m_cov` and `y_cov` are ignored by the procedure. |
| `interaction` | A logical variable specifying whether there is an exposure-mediator interaction term in the outcome model. |
| `Firth` | A logical variable specifying whether to compute conventional or penalized maximum likelihood estimates for the logistic regression models (see details). |
| `boot` | A logical value specifying whether the confidence intervals are obtained by the delta method or by percentile bootstrap. |
| `nboot` | The number of bootstrap replications used to obtain the confidence intervals if `boot = TRUE`. By default `nboot = 1000`. |
| `bootseed` | The value of the initial seed (positive integer) for random number generation if `boot = TRUE`. |
| `confcoef` | A number between 0 and 1 for the confidence coefficient (ex.: 0.95) of the interval estimates. |
| `hvalue_m` | The value corresponding to the high level of the mediator. If the mediator is already coded as a numerical binary variable taking 0 or 1 values, then by default `hvalue_m = 1`. |
| `hvalue_y` | The value corresponding to the high level of the outcome. If the outcome is already coded as a numerical binary variable taking 0 or 1 values, then by default `hvalue_y = 1`. |
| `yprevalence` | The prevalence of the outcome in the population (a number between 0 and 1). Option used when case-control data are used. The low level of the outcome is treated as the control level. |

## Details

By default, `exactmed()` reports mediation effects evaluated at the sample-specific mean values of the numerical covariates (including the dummy variables created internally by the function to represent the non-reference levels of the categorical covariates). In order to estimate mediation effects at specific values of some covariates (that is, stratum-specific effects), the user needs to provide named vectors `m_cov_cond` and/or `y_cov_cond` containing those values or levels. The adjustment covariates appearing in both `m_cov` and `y_cov` (common adjustment covariates) must have the same values; otherwise, `exactmed()`'s execution is aborted and an error message is displayed in the R console.

The Firth parameter allows to reduce the bias of the regression coefficients estimators when facing a problem of separation or quasi-separation. The bias reduction is achieved by the `brglmFit` fitting method of the *brglm2* package. More precisely, estimates are obtained via penalized maximum likelihood with a Jeffreys prior penalty, which is equivalent to the mean bias-reducing adjusted score equation approach in Firth (1993).

When the data come from a case-control study, the `yprevalence` parameter should be used and its value ideally correspond to the true outcome prevalence. `exactmed()` accounts for the ascertainment in the sample by employing weighted regression techniques that use inverse probability weighting (IPW) with robust standard errors. These errors are obtained via the `vcovHC` function of the R package *sandwich*. Specifically, we use the HC3 type covariance matrix estimator (default type of the `vcovHC` function).

For the mediation effects expressed on the multiplicative scales (odds ratio, OR; risk ratio, RR), the `exactmed()` function returns delta method confidence intervals by exponentiating the lower and upper limits of the normal confidence intervals obtained for the logarithmic transformations of the effects. The `exactmed()` function also provides the estimated standard errors of natural and controlled direct effects estimators that are not log-transformed, where those are derived using a first order Taylor expansion (e.g., $\hat{SE}(\hat{OR}) = \hat{OR} \times \hat{SE}(\log(\hat{OR}))$). The function performs Z-tests (null hypothesis: there is no effect) computing the corresponding two-tailed *p*-values. Note that for the multiplicative scales, the standard scores (test statistics) are obtained by dividing the logarithm of an effect estimator by the estimator of the corresponding standard error (e.g., $\log(\hat{OR})/\hat{SE}(\log(\hat{OR}))$). No log-transformation is applied when working on the risk difference scale.

## Value

An object of class `results` is returned:

| | |
|---|---|
| `ne.or` | Natural effects estimates on OR scale. |
| `ne.rr` | Natural effects estimates on RR scale. |
| `ne.rd` | Natural effects estimates on RD scale. |
| `cdem0` | Controlled direct effect (m=0) estimates. |
| `cdem1` | Controlled direct effect (m=1) estimates. |
| `med.reg` | Summary of the mediator regression. |
| `out.reg` | Summary of the outcome regression. |

If `boot==TRUE`, the returned object also contains:

| | |
|---|---|
| `boot.ne.or` | Bootstrap replications of natural effects on OR scale. |
| `boot.ne.rr` | Bootstrap replications of natural effects on RR scale. |
| `boot.ne.rd` | Bootstrap replications of natural effects on RD scale. |
| `boot.cdem0.or` | Bootstrap replications of controlled direct effect (m=0) on OR scale. |

| | |
|---|---|
| `boot.cdem0.rr` | Bootstrap replications of controlled direct effect (m=0) on RR scale. |
| `boot.cdem0.rd` | Bootstrap replications of controlled direct effect (m=0) on RD scale. |
| `boot.cdem1.or` | Bootstrap replications of controlled direct effect (m=1) on OR scale. |
| `boot.cdem1.rr` | Bootstrap replications of controlled direct effect (m=1) on RR scale. |
| `boot.cdem1.rd` | Bootstrap replications of controlled direct effect (m=1) on RD scale. |
| `boot.ind` | Indices of the observations sampled in each bootstrap replication (one replication per column). |

## Note

The `exactmed()` function only works for complete data. Users can apply multiple imputation techniques (e.g., R package *mice*) or remove observations of variables used in mediation analysis that have missing values (NA).

## References

Samoilenko M, Blais L, Lefebvre G. Comparing logistic and log-binomial models for causal mediation analyses of binary mediators and rare binary outcomes: evidence to support cross-checking of mediation results in practice. *Observational Studies*.2018;4(1):193-216. doi:10.1353/obs.2018.0013.

Samoilenko M, Lefebvre G. Parametric-regression-based causal mediation analysis of binary outcomes and binary mediators: moving beyond the rareness or commonness of the outcome. *American Journal of Epidemiology*.2021;190(9):1846-1858. doi:10.1093/aje/kwab055.

Samoilenko M, Lefebvre G. An exact regression-based approach for the estimation of the natural direct and indirect effects with a binary outcome and a continuous mediator. *Statistics in Medicine*.2023; 42(3): 353–387. doi:10.1002/sim.9621.

Firth D. Bias reduction of maximum likelihood estimates. *Biometrika*.1993;80:27-38. doi:10.2307/2336755.

## Examples

```
exactmed(
  data = datamed, a = "X", m = "M", y = "Y", a1 = 1, a0 = 0,
  m_cov = c("C1", "C2"), y_cov = c("C1", "C2")
)

exactmed(
  data = datamed, a = "X", m = "M", y = "Y", a1 = 1, a0 = 0,
  m_cov = c("C1", "C2"), y_cov = c("C1", "C2"), yprevalence = 0.1
)

m_cov_cond <- c(C1 = 0.1, C2 = 0.4)
y_cov_cond <- c(C1 = 0.1, C2 = 0.4)

exactmed(
  data = datamed, a = "X", m = "M", y = "Y", a1 = 1, a0 = 0,
  m_cov = c("C1", "C2"), y_cov = c("C1", "C2"),
  m_cov_cond = m_cov_cond, y_cov_cond = y_cov_cond
)
```

```
C1b <- factor(sample(c("a", "b", "c"), nrow(datamed), replace = TRUE))
datamed$C1 <- C1b

m_cov_cond <- list(C1 = "c", C2 = 0.4)
y_cov_cond <- list(C1 = "c", C2 = 0.4)

exactmed(
  data = datamed, a = "X", m = "M", y = "Y", a1 = 1, a0 = 0,
  m_cov = c("C1", "C2"), y_cov = c("C1", "C2"),
  m_cov_cond = m_cov_cond, y_cov_cond = y_cov_cond
)
```

---

## `exactmed_c` *Exact Mediation Effects Computation (Continuous Mediator)*

### Description

Relying on a regression-based approach, the `exactmed_c()` function calculates standard causal mediation effects when the outcome is binary and the mediator is continuous. More precisely, `exactmed_c()` relies on logistic and linear models for the outcome and mediator, respectively, in order to compute *exact* conditional natural direct and indirect effects. Nested counterfactual probabilities underlying the definition of natural effects are calculated using numerical integration. The function returns point and interval estimates for the conditional natural effects without making any assumption regarding the rareness or commonness of the outcome (hence the term exact). For completeness, `exactmed_c()` also calculates the conditional controlled direct effect at a specified value of the mediator. Natural and controlled effects estimates are reported using three different scales: odds ratio (OR), risk ratio (RR) and risk difference (RD). The interval estimates can be obtained either by the delta method or the bootstrap.

### Usage

```
exactmed_c(
  data,
  a,
  m,
  y,
  a1,
  a0,
  m_cov = NULL,
  y_cov = NULL,
  m_cov_cond = NULL,
  y_cov_cond = NULL,
  adjusted = TRUE,
  interaction = TRUE,
  Firth = FALSE,
  boot = FALSE,
  nboot = 1000,
```

```
  bootseed = 1991,
  confcoef = 0.95,
  hvalue_y = NULL,
  yprevalence = NULL,
  mf = NULL
)
```

### Arguments

| | |
|---|---|
| `data` | A named data frame that includes the exposure, mediator and outcome variables as well as the covariates to be adjusted for in the models. The exposure can be either binary or continuous. If a covariate is categorical, it has to be included in the data frame as a factor, character or logical variable. |
| `a` | The name of the binary or continuous exposure variable. |
| `m` | The name of the continuous mediator variable. |
| `y` | The name of the binary outcome variable. |
| `a1` | A value corresponding to the high level of the exposure. |
| `a0` | A value corresponding to the low level of the exposure. |
| `m_cov` | A vector containing the names of the adjustment variables (covariates) in the mediator model. |
| `y_cov` | A vector containing the names of the adjustment variables (covariates) in the outcome model. |
| `m_cov_cond` | A named vector (atomic vector or list) containing specific values for some or all of the adjustment covariates `m_cov` in the mediator model. Please consult the package vignette for details. |
| `y_cov_cond` | A named vector (atomic vector or list) containing specific values for some or all of the adjustment covariates `y_cov` in the outcome model. Please consult the package vignette for details. |
| `adjusted` | A logical variable specifying whether to obtain adjusted or unadjusted estimates. If `adjusted = FALSE`, vectors `m_cov` and `y_cov` are ignored by the procedure. |
| `interaction` | A logical variable specifying whether there is an exposure-mediator interaction term in the outcome model. |
| `Firth` | A logical variable specifying whether to compute conventional or penalized maximum likelihood estimates for the outcome logistic regression model. |
| `boot` | A logical value specifying whether the confidence intervals are obtained by the delta method or by percentile bootstrap. |
| `nboot` | The number of bootstrap replications used to obtain the confidence intervals if `boot = TRUE`. By default `nboot = 1000`. |
| `bootseed` | The value of the initial seed (positive integer) for random number generation if `boot = TRUE`. |
| `confcoef` | A number between 0 and 1 for the confidence coefficient (ex.: 0.95) of the interval estimates. |

| | |
|---|---|
| `hvalue_y` | The value corresponding to the high level of the outcome. If the outcome is already coded as a numerical binary variable taking 0 or 1 values, then by default `hvalue_y = 1`. |
| `yprevalence` | The prevalence of the outcome in the population (a number between 0 and 1). Option used when case-control data are used. The low level of the outcome is treated as the control level. |
| `mf` | The value of the mediator at which the conditional controlled direct effect is computed. If it is not specified, `mf` is fixed at the sample-specific mean of the mediator (default). |

## Details

By default, `exactmed_c()` reports mediation effects evaluated at the sample-specific mean values of the numerical covariates (including the dummy variables created internally by the function to represent the non-reference levels of the categorical covariates). In order to estimate mediation effects at specific values of some covariates (that is, stratum-specific effects), the user needs to provide named vectors `m_cov_cond` and/or `y_cov_cond` containing those values or levels. The adjustment covariates appearing in both `m_cov` and `y_cov` (common adjustment covariates) must have the same values; otherwise, `exactmed_c()`'s execution is aborted and an error message is displayed in the R console.

The Firth parameter allows to reduce the bias of the outcome logistic regression coefficients estimators when facing a problem of separation or quasi-separation. The bias reduction is achieved by the `brglmFit` fitting method of the *brglm2* package. More precisely, estimates are obtained via penalized maximum likelihood with a Jeffreys prior penalty, which is equivalent to the mean bias-reducing adjusted score equation approach in Firth (1993).

When the data come from a case-control study, the `yprevalence` parameter should be used and its value ideally correspond to the true outcome prevalence. `exactmed_c()` accounts for the ascertainment in the sample by employing weighted regression techniques that use inverse probability weighting (IPW) with robust standard errors. These errors are obtained via the `vcovHC` function of the R package *sandwich*. Specifically, we use the HC3 type covariance matrix estimator (default type of the `vcovHC` function).

For the mediation effects expressed on the multiplicative scales (odds ratio, OR; risk ratio, RR), the `exactmed_c()` function returns delta method confidence intervals by exponentiating the lower and upper limits of the normal confidence intervals obtained for the logarithmic transformations of the effects. The `exactmed_c()` function also provides the estimated standard errors of natural and controlled direct effects estimators that are not log-transformed, where those are derived using a first order Taylor expansion (e.g., $\hat{SE}(\hat{OR}) = \hat{OR} \times \hat{SE}(\log(\hat{OR}))$). The function performs Z-tests (null hypothesis: there is no effect) computing the corresponding two-tailed *p*-values. Note that for the multiplicative scales, the standard scores (test statistics) are obtained by dividing the logarithm of an effect estimator by the estimator of the corresponding standard error (e.g., $\log(\hat{OR})/\hat{SE}(\log(\hat{OR}))$). No log-transformation is applied when working on the risk difference scale.

## Value

An object of class `results_c` is returned:

| | |
|---|---|
| `ne.or` | Natural effects estimates on OR scale. |

| | |
|---|---|
| `ne.rr` | Natural effects estimates on RR scale. |
| `ne.rd` | Natural effects estimates on RD scale. |
| `cde` | Controlled direct effect estimates. |
| `med.reg` | Summary of the mediator regression. |
| `out.reg` | Summary of the outcome regression. |

If `boot==TRUE`, the returned object also contains:

| | |
|---|---|
| `boot.ne.or` | Bootstrap replications of natural effects on OR scale. |
| `boot.ne.rr` | Bootstrap replications of natural effects on RR scale. |
| `boot.ne.rd` | Bootstrap replications of natural effects on RD scale. |
| `boot.cde.or` | Bootstrap replications of controlled direct effect on OR scale. |
| `boot.cde.rr` | Bootstrap replications of controlled direct effect on RR scale. |
| `boot.cde.rd` | Bootstrap replications of controlled direct effect on RD scale. |
| `boot.ind` | Indices of the observations sampled in each bootstrap replication (one replication per column). |

## Note

The `exactmed_c()` function only works for complete data. Users can apply multiple imputation techniques (e.g., R package *mice*) or remove observations of variables used in mediation analysis that have missing values (NA).

## References

Samoilenko M, Blais L, Lefebvre G. Comparing logistic and log-binomial models for causal mediation analyses of binary mediators and rare binary outcomes: evidence to support cross-checking of mediation results in practice. *Observational Studies*.2018;4(1):193-216.

Samoilenko M, Lefebvre G. Parametric-regression-based causal mediation analysis of binary outcomes and binary mediators: moving beyond the rareness or commonness of the outcome. *American Journal of Epidemiology*.2021;190(9):1846-1858. doi:10.1093/aje/kwab055.

Samoilenko M, Lefebvre G. An exact regression-based approach for the estimation of the natural direct and indirect effects with a binary outcome and a continuous mediator. *Statistics in Medicine*.2023; 42(3): 353–387. doi:10.1002/sim.9621.

Firth D. Bias reduction of maximum likelihood estimates. *Biometrika*.1993;80:27-38. doi:10.2307/2336755.

## Examples

```
exactmed_c(
  data = datamed_c, a = "X", m = "M", y = "Y", a1 = 1, a0 = 0,
  m_cov = c("C1", "C2"), y_cov = c("C1", "C2")
)

exactmed_c(
  data = datamed_c, a = "X", m = "M", y = "Y", a1 = 1, a0 = 0,
```

```
  m_cov = c("C1", "C2"), y_cov = c("C1", "C2"), yprevalence = 0.1
)

m_cov_cond <- c(C1 = 0.1, C2 = 0.4)
y_cov_cond <- c(C1 = 0.1, C2 = 0.4)

exactmed_c(
  data = datamed_c, a = "X", m = "M", y = "Y", a1 = 1, a0 = 0,
  m_cov = c("C1", "C2"), y_cov = c("C1", "C2"),
  m_cov_cond = m_cov_cond, y_cov_cond = y_cov_cond
)

C1b <- factor(sample(c("a", "b", "c"), nrow(datamed_c), replace = TRUE))
datamed_c$C1 <- C1b

m_cov_cond <- list(C1 = "c", C2 = 0.4)
y_cov_cond <- list(C1 = "c", C2 = 0.4)

exactmed_c(
  data = datamed_c, a = "X", m = "M", y = "Y", a1 = 1, a0 = 0,
  m_cov = c("C1", "C2"), y_cov = c("C1", "C2"),
  m_cov_cond = m_cov_cond, y_cov_cond = y_cov_cond
)
```

---

## `exactmed_cat` *Exact Mediation Effects Computation (Categorical Mediator)*

### Description

Relying on a regression-based approach, the `exactmed_cat()` function calculates standard causal mediation effects when the outcome is binary and the mediator is categorical. More precisely, `exactmed_cat()` relies on binary and multinomial logistic regression models for the outcome and mediator, respectively, in order to compute *exact* conditional natural direct and indirect effects. The function returns point and interval estimates for the conditional natural effects without making any assumption regarding the rareness or commonness of the outcome (hence the term exact). For completeness, `exactmed_cat()` also calculates the conditional controlled direct effect at a specified level of the mediator. Natural and controlled effects estimates are reported using three different scales: odds ratio (OR), risk ratio (RR) and risk difference (RD). The interval estimates can be obtained either by the delta method or the bootstrap.

### Usage

```
exactmed_cat(
  data,
  a,
  m,
  y,
  a1,
  a0,
```

```
    m_cov = NULL,
    y_cov = NULL,
    m_cov_cond = NULL,
    y_cov_cond = NULL,
    adjusted = TRUE,
    interaction = TRUE,
    Firth = FALSE,
    boot = FALSE,
    nboot = 1000,
    bootseed = 1991,
    confcoef = 0.95,
    blevel_m = NULL,
    hvalue_y = NULL,
    yprevalence = NULL,
    mf = NULL
  )
```

## Arguments

| | |
|---|---|
| `data` | A named data frame that includes the exposure, mediator and outcome variables as well as the covariates to be adjusted for in the models. The exposure can be either binary or continuous. If a covariate is categorical, it has to be included in the data frame as a factor, character or logical variable. |
| `a` | The name of the binary or continuous exposure variable. |
| `m` | The name of the categorical mediator variable. |
| `y` | The name of the binary outcome variable. |
| `a1` | A value corresponding to the high level of the exposure. |
| `a0` | A value corresponding to the low level of the exposure. |
| `m_cov` | A vector containing the names of the adjustment variables (covariates) in the mediator model. |
| `y_cov` | A vector containing the names of the adjustment variables (covariates) in the outcome model. |
| `m_cov_cond` | A named vector (atomic vector or list) containing specific values for some or all of the adjustment covariates `m_cov` in the mediator model. Please consult the package vignette for details. |
| `y_cov_cond` | A named vector (atomic vector or list) containing specific values for some or all of the adjustment covariates `y_cov` in the outcome model. Please consult the package vignette for details. |
| `adjusted` | A logical variable specifying whether to obtain adjusted or unadjusted estimates. If `adjusted = FALSE`, vectors `m_cov` and `y_cov` are ignored by the procedure. |
| `interaction` | A logical variable specifying whether there are exposure-mediator interaction terms in the outcome model. |
| `Firth` | A logical variable specifying whether to compute conventional or penalized maximum likelihood estimates for the outcome logistic regression model. |

| | |
|---|---|
| `boot` | A logical value specifying whether the confidence intervals are obtained by the delta method or by percentile bootstrap. |
| `nboot` | The number of bootstrap replications used to obtain the confidence intervals if `boot = TRUE`. By default `nboot = 1000`. |
| `bootseed` | The value of the initial seed (positive integer) for random number generation if `boot = TRUE`. |
| `confcoef` | A number between 0 and 1 for the confidence coefficient (ex.: 0.95) of the interval estimates. |
| `blevel_m` | The reference level of the mediator. If it is not specified `blevel_m` is fixed to the first level of the mediator (default). |
| `hvalue_y` | The value corresponding to the high level of the outcome. If the outcome is already coded as a numerical binary variable taking 0 or 1 values, then by default `hvalue_y = 1`. |
| `yprevalence` | The prevalence of the outcome in the population (a number between 0 and 1). Option used when case-control data are used. The low level of the outcome is treated as the control level. |
| `mf` | The level of the mediator at which the conditional controlled direct effect is computed. If it is not specified, `mf` is fixed at the reference level (`blevel_m`) of the mediator (default). |

## Details

By default, `exactmed_cat()` reports mediation effects evaluated at the sample-specific mean values of the numerical covariates (including the dummy variables created internally by the function to represent the non-reference levels of the categorical covariates). In order to estimate mediation effects at specific values of some covariates (that is, stratum-specific effects), the user needs to provide named vectors `m_cov_cond` and/or `y_cov_cond` containing those values or levels. The adjustment covariates appearing in both `m_cov` and `y_cov` (common adjustment covariates) must have the same values; otherwise, `exactmed_cat()`'s execution is aborted and an error message is displayed in the R console.

The Firth parameter allows to reduce the bias of the outcome logistic regression coefficients estimators when facing a problem of separation or quasi-separation. The bias reduction is achieved by the `brglmFit` fitting method of the *brglm2* package. More precisely, estimates are obtained via penalized maximum likelihood with a Jeffreys prior penalty, which is equivalent to the mean bias-reducing adjusted score equation approach in Firth (1993).

When the data come from a case-control study, the `yprevalence` parameter should be used and its value ideally correspond to the true outcome prevalence. `exactmed_cat()` accounts for the ascertainment in the sample by employing weighted regression techniques that use inverse probability weighting (IPW) with robust standard errors. These errors are obtained via the `vcovHC` function of the R package *sandwich*. Specifically, we use the HC3 type covariance matrix estimator (default type of the `vcovHC` function).

For the mediation effects expressed on the multiplicative scales (odds ratio, OR; risk ratio, RR), the `exactmed_cat()` function returns delta method confidence intervals by exponentiating the lower and upper limits of the normal confidence intervals obtained for the logarithmic transformations of the effects. The `exactmed_cat()` function also provides the estimated standard errors of natural and controlled direct effects estimators that are not log-transformed, where those are derived using a first order Taylor expansion (e.g., $\hat{SE}(\hat{OR}) = \hat{OR} \times \hat{SE}(\log(\hat{OR}))$). The function performs Z-tests (null hypothesis: there is no effect) computing the corresponding two-tailed *p*-values. Note that for the multiplicative scales, the standard scores (test statistics) are obtained by dividing the logarithm of an effect estimator by the estimator of the corresponding standard error (e.g., $\log(\hat{OR})/\hat{SE}(\log(\hat{OR}))$). No log-transformation is applied when working on the risk difference scale.

## Value

An object of class `results_cat` is returned:

| | |
|---|---|
| `ne.or` | Natural effects estimates on OR scale. |
| `ne.rr` | Natural effects estimates on RR scale. |
| `ne.rd` | Natural effects estimates on RD scale. |
| `cde` | Controlled direct effect estimates. |
| `med.reg` | Summary of the mediator regression. |
| `out.reg` | Summary of the outcome regression. |

If `boot==TRUE`, the returned object also contains:

| | |
|---|---|
| `boot.ne.or` | Bootstrap replications of natural effects on OR scale. |
| `boot.ne.rr` | Bootstrap replications of natural effects on RR scale. |
| `boot.ne.rd` | Bootstrap replications of natural effects on RD scale. |
| `boot.cde.or` | Bootstrap replications of controlled direct effect on OR scale. |
| `boot.cde.rr` | Bootstrap replications of controlled direct effect on RR scale. |
| `boot.cde.rd` | Bootstrap replications of controlled direct effect on RD scale. |
| `boot.ind` | Indices of the observations sampled in each bootstrap replication (one replication per column). |

## Note

The `exactmed_cat()` function only works for complete data. Users can apply multiple imputation techniques (e.g., R package *mice*) or remove observations of variables used in mediation analysis that have missing values (NA).

## References

Samoilenko M, Blais L, Lefebvre G. Comparing logistic and log-binomial models for causal mediation analyses of binary mediators and rare binary outcomes: evidence to support cross-checking of mediation results in practice. *Observational Studies*.2018;4(1):193-216.

Samoilenko M, Lefebvre G. Parametric-regression-based causal mediation analysis of binary outcomes and binary mediators: moving beyond the rareness or commonness of the outcome. *American Journal of Epidemiology*.2021;190(9):1846-1858. doi:10.1093/aje/kwab055.

Samoilenko M, Lefebvre G. An exact regression-based approach for the estimation of the natural direct and indirect effects with a binary outcome and a continuous mediator. *Statistics in Medicine*.2023; 42(3): 353–387. doi:10.1002/sim.9621.

Firth D. Bias reduction of maximum likelihood estimates. *Biometrika*.1993;80:27-38. doi:10.2307/2336755.

## Examples

```
exactmed_cat(
  data = datamed_cat, a = "X", m = "M", y = "Y", a1 = 1, a0 = 0,
  m_cov = c("C1", "C2"), y_cov = c("C1", "C2")
)

exactmed_cat(
  data = datamed_cat, a = "X", m = "M", y = "Y", a1 = 1, a0 = 0,
  m_cov = c("C1", "C2"), y_cov = c("C1", "C2"), yprevalence = 0.1
)

m_cov_cond <- c(C1 = 0.1, C2 = 0.4)
y_cov_cond <- c(C1 = 0.1, C2 = 0.4)

exactmed_cat(
  data = datamed_cat, a = "X", m = "M", y = "Y", a1 = 1, a0 = 0,
  m_cov = c("C1", "C2"), y_cov = c("C1", "C2"),
  m_cov_cond = m_cov_cond, y_cov_cond = y_cov_cond
)

C1b <- factor(sample(c("a", "b", "c"), nrow(datamed_cat), replace = TRUE))
datamed_cat$C1 <- C1b

m_cov_cond <- list(C1 = "c", C2 = 0.4)
y_cov_cond <- list(C1 = "c", C2 = 0.4)

exactmed_cat(
  data = datamed_cat, a = "X", m = "M", y = "Y", a1 = 1, a0 = 0,
  m_cov = c("C1", "C2"), y_cov = c("C1", "C2"),
  m_cov_cond = m_cov_cond, y_cov_cond = y_cov_cond
)
```

# Index