Package ‘rqlm’

January 28, 2026

Type Package

Title Modified Poisson Regression for Binary Outcome and Related Methods

Version 4.3-1

Date 2026-01-28

Maintainer Hisashi Noma <noma@ism.ac.jp>

Description Modified Poisson, logistic and least-squares regression analyses for binary outcomes of Zou (2004) <doi:10.1093/aje/kwh090>, Noma (2025)<Forthcoming>, and Cheung (2007) <doi:10.1093/aje/kwm223> have been standard multivariate analysis methods to estimate risk ratio and risk difference in clinical and epidemiological studies. This R package involves an easy-to-handle function to implement these analyses by simple commands. Missing data analysis tools (multiple imputation) are also involved. In addition, recent studies have shown the ordinary robust variance estimator possibly has serious bias under small or moderate sample size situations for these methods. This package also provides computational tools to calculate alternative accurate confidence intervals.

Depends R (>= 3.5.0)

Imports stats, MASS, nnet, sandwich, mice, lme4

License GPL-3

Encoding UTF-8

LazyData true

NeedsCompilation no

Author Hisashi Noma [aut, cre] (ORCID: <https://orcid.org/0000-0002-2520-9949>)

Repository CRAN

Date/Publication 2026-01-28 08:50:02 UTC

Contents

`rqlm-package` *The 'rqlm' package.*

Description

Modified Poisson, logistic and least-squares regression analyses for binary outcomes have been standard multivariate analysis methods to estimate risk ratio and risk difference in clinical and epidemiological studies. This R package involves an easy-to-handle function to implement these analyses by simple commands. Missing data analysis tools (multiple imputation) are also involved. In addition, recent studies have shown the ordinary robust variance estimator possibly has serious bias under small or moderate sample size situations for these methods. This package also provides computational tools to calculate accurate confidence intervals. Also, standard computational tools for target trial emulation are included.

References

Cheung, Y. B. (2007). A modified least-squares regression approach to the estimation of risk difference. *American Journal of Epidemiology* **166**, 1337-1344.

Hernan, M. A., Wang, W., and Leaf, D. E. (2022). Target Trial Emulation: A Framework for Causal Inference From Observational Data. *JAMA* **328**, 2446-2447.

Noma, H. (2025). Robust variance estimators for risk ratio estimators from logistic regression in cohort and case-cohort studies. Forthcoming.

Noma, H. and Gosho, M. (2025). Finite-sample improved confidence intervals based on the estimating equation theory for the modified Poisson and least-squares regressions. *Epidemiologic Methods* **14**, 20240030.

Noma, H. and Gosho, M. (2025). Logistic mixed-effects model analysis with pseudo-observations for estimating risk ratios in clustered binary data analysis. *Statistics in Medicine* **44**, e70280.

Noma, H., Sunada, H., and Gosho, M. (2025). Quasi-likelihood ratio tests and the Bartlett-type correction for improved inferences of the modified Poisson and least-squares regressions for binary outcomes. *Statistica Neerlandica* **79**, e70012.

Shiiba, H. and Noma, H. (2025). Confidence intervals of risk ratios for the augmented logistic regression with pseudo-observations. *Stats* **8**, 83.

Zou, G. (2004). A modified poisson regression approach to prospective studies with binary data. *American Journal of Epidemiology* **159**, 702-706.

`bsci.ls` *Calculating bootstrap confidence interval for modified least-squares regression based on the quasi-score statistic*

Description

Recent studies revealed the robust standard error estimates of the modified least-squares regression analysis are generally biased under small or moderate sample settings. To adjust the bias and to provide more accurate confidence intervals, confidence interval and P-value of the test for risk difference by modified least-squares regression are calculated based on the bootstrap approach of Noma and Gosho (2024).

Usage

```
bsci.ls(formula, data, x.name=NULL, B=1000, cl=0.95, C0=10^-5,
 digits=4, seed=527916)
```

Arguments

`formula`	An object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be fitted.
`data`	A data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the variables in the model.
`x.name`	The variable name that the confidence interval is calculated for the regression coefficient; should be involved in `formula` as an explanatory variable. Specify as a character object.
`B`	The number of bootstrap resampling (default: 1000)
`cl`	Confidence level for calculating confidence intervals (default: 0.95)
`C0`	A tuning parameter to control the precisions of numerical computations of confidence limits (default: 10^-5).
`digits`	Number of decimal places in the output (default: 4).
`seed`	Seed to generate random numbers (default: 527916).

Value

Results of the modified least-squares analyses are presented. Three objects are provided: Results of the modified least-squares regression with the Wald-type approximation by `rqlm`, the bootstrap-based confidence interval for the corresponding covariate, and P-value for the bootstrap test of `RD=0`.

References

Noma, H. and Gosho, M. (2025). Finite-sample improved confidence intervals based on the estimating equation theory for the modified Poisson and least-squares regressions. *Epidemiologic Methods* **14**, 20240030.

Examples

```
data(exdata01)

bsci.ls(y ~ x1 + x2 + x3 + x4, data=exdata01, "x3", B=10)
# For illustration. B should be >= 1000 (the number of bootstrap resampling).
```

`bsci.pois` *Calculating bootstrap confidence interval for modified Poisson regression based on the quasi-score statistic*

Description

Recent studies revealed the risk ratio estimates and robust standard error estimates of the modified Poisson regression analysis are generally biased under small or moderate sample settings. To adjust the bias and to provide more accurate confidence intervals, confidence interval and P-value of the test for risk ratio by modified Poisson regression are calculated based on the bootstrap approach of Noma and Gosho (2024).

Usage

```
bsci.pois(formula, data, x.name=NULL, B=1000, eform=FALSE, cl=0.95, C0=10^-5,
 digits=4, seed=527916)
```

Arguments

`formula`	An object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be fitted.
`data`	A data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the variables in the model.
`x.name`	The variable name that the confidence interval is calculated for the regression coefficient; should be involved in `formula` as an explanatory variable. Specify as a character object.
`B`	The number of bootstrap resampling (default: 1000)
`eform`	A logical value that specify whether the outcome should be transformed by exponential function (default: `FALSE`)
`cl`	Confidence level for calculating confidence intervals (default: 0.95)
`C0`	A tuning parameter to control the precisions of numerical computations of confidence limits (default: 10^-5).
`digits`	Number of decimal places in the output (default: 4).
`seed`	Seed to generate random numbers (default: 527916).

Value

Results of the modified Poisson analyses are presented. Three objects are provided: Results of the modified Poisson regression with the Wald-type approximation by `rqlm`, the bootstrap confidence interval for the corresponding covariate, and P-value for the bootstrap test of `RR=1`.

References

Noma, H. and Gosho, M. (2025). Finite-sample improved confidence intervals based on the estimating equation theory for the modified Poisson and least-squares regressions. *Epidemiologic Methods* **14**, 20240030.

Examples

```
data(exdata01)

bsci.pois(y ~ x1 + x2 + x3 + x4, data=exdata01, "x3", B=10, eform=TRUE)
# For illustration. B should be >= 1000 (the number of bootstrap resampling).
```

---

coeff — *Computation of the ordinary confidence intervals and P-values using the model variance estimator*

Description

Confidence intervals and P-values for the generalized linear model and generalized linear-mixed-effects model can be calculated using the ordinary model variance estimators. Through simply entering the output objects of `lm`, `glm`, `lmer`, or `glmer`, the inference results are fastly computed. For the linear regression model, the exact confidence intervals and P-values based on the t-distribution are calculated. Also, for the generalized linear model, the Wald-type confidence intervals and P-values based on the asymptotic normal approximation are computed. The resultant coefficients and confidence limits can be transformed to exponential scales by specifying `eform`.

Usage

```
coeff(gm, eform=FALSE, cl=0.95, digits=4)
```

Arguments

`gm`	An output object of `lm`, `glm`, `lmer`, or `glmer`.
`eform`	A logical value that specify whether the outcome should be transformed by exponential function (default: `FALSE`)
`cl`	Confidence level for calculating confidence intervals (default: 0.95)
`digits`	Number of decimal places in the output (default: 4).

Value

Results of inferences of the regression coefficients using the ordinary model variance estimators.

- `coef`: Coefficient estimates; transformed to the exponential scale if `eform=TRUE`.
- `SE`: Robust standard error estimates for `coef`.
- `CL`: Lower limits of confidence intervals.
- `CU`: Upper limits of confidence intervals.
- `P-value`: P-values for the coefficient tests.

Examples

```
data(exdata02)

gm1 <- glm(y ~ x1 + x2 + x3 + x4, data=exdata02, family=binomial)
coeff(gm1,eform=TRUE)
# Logistic regression analysis
# Coefficient estimates are translated to odds ratio scales

lm1 <- lm(x1 ~ x2 + x3 + x4, data=exdata02)
coeff(lm1)
# Linear regression analysis

data(mch)

if(requireNamespace("lme4", quietly = TRUE)) {
    lmr1 <- lme4::lmer(Y ~ x + (1|SOUM), data=mch)
    coeff(lmr1)
}
# Linear mixed-effects model analysis

if(requireNamespace("lme4", quietly = TRUE)) {
gmr1 <- lme4::glmer(y ~ x + (1|SOUM), nAGQ=25, family=binomial, data=mch)
coeff(gmr1, eform=TRUE)
}
# Logistic mixed-effects model analysis
# Coefficient estimates are translated to odds ratio scales
```

`exdata01` *A simulated example dataset*

Description

A simulated cohort data with binomial outcome.

- `y`: Dichotomous outcome variable.
- `x1`: Continuous covariate.
- `x2`: Binary covariate.
- `x3`: Binary covariate.
- `x4`: Binary covariate.

Usage

```
data(exdata01)
```

Format

A simulated cohort data with binomial outcome (n=40).

`exdata02` *A simulated example dataset*

Description

A simulated cohort data with binomial outcome.

- `y`: Dichotomous outcome variable.
- `x1`: Continuous covariate.
- `x2`: Binary covariate.
- `x3`: Binary covariate.
- `x4`: Binary covariate.

Usage

```
data(exdata02)
```

Format

A simulated cohort data with binomial outcome (n=1200).

`exdata03` *A simulated example dataset with missing covariates*

Description

A simulated cohort data with binomial outcome. Some covariates involve missing data.

- `y`: Dichotomous outcome variable.
- `x1`: Continuous covariate.
- `x2`: Binary covariate.
- `x3`: Binary covariate.
- `x4`: Binary covariate.

Usage

```
data(exdata03)
```

Format

A simulated cohort data with binomial outcome (n=1200). Some covariates involve missing data.

`exdata04` *A simulated example dataset for target trial emulation*

Description

A simulated data for sequentially nested emulated trials.

- `ID`: The id variable for individual participants.
- `trial`: The trial id variable.
- `time`: The discrete time index within each emulated trial, representing the interval since trial baseline.
- `A`: Treatment variable.
- `L1`: Baseline confounding variable.
- `Y`: Dichotomous outcome variable.
- `w_pp`: The weight variable.

Usage

```
data(exdata04)
```

Format

A simulated data for sequentially nested emulated trials (n=600).

`mch` *A cluster-randomised trial dataset for the maternal and child health handbook*

Description

A cluster-randomised trial dataset with binomial outcome.

- `ID`: ID variable of participants.
- `SOUM`: ID variable of soums (involving 18 soums).
- `x`: Binary variable specifying intervention groups (1=Intervention, 0=Control).
- `mage`: Mother's age.
- `medu`: Mother's education (1=uneducated, 2=elementary, 3=incomplete secondary, 4=complete secondary, 5=incomplete high, 6=high (completed collage or university)).

- mmarry: Mother's marital status (1=single, 2=married/cohabitating, 3=separated/divorce, 4=windowed/other).
- mprig1: First pregnancy (1=Yes, 2=No).
- height: Mother's height.
- weight: Mother's weight.
- time: Travel time from mother's home to antenatal care clinic.
- Y: Outcome variable: Number of antenatal visits.
- y: Outcome variable: Whether the number of antenatal visits is >= 6 (0 or 1).
- ses: Quintile groups by the social-economic index (= 1, 2, 3, 4, 5).

Usage

```
data(mch)
```

Format

A data frame with 500 participants with 18 soums.

References

Mori, R., Yonemoto, N., Noma, H., et al. (2015). The Maternal and Child Health (MCH) handbook in Mongolia: a cluster-randomized, controlled trial. *PloS One* **10**: e0119772.

---

mi_glm — *Multiple imputation analysis for the generalized linear model*

Description

Multiple imputation analysis for the generalized linear model is performed for the imputed datasets generated by mice function in mice package. For computing covariance matrix estimate, the ordinary Rubin's rule is adapted to the model variance estimates.

Usage

```
mi_glm(ice, formula, family=gaussian, offset=NULL, eform=FALSE, cl=0.95, digits=4)
```

Arguments

ice	An output object of mice function in mice package.
formula	An object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be fitted.
family	A description of the error distribution and link function to be used in the model.
offset	A vector of offset. This can be used to specify an a priori known component to be included in the linear predictor during fitting. This should be NULL or a numeric vector of length equal to the number of cases.

eform	A logical value that specify whether the outcome should be transformed by exponential function (default: FALSE)
cl	Confidence level for calculating confidence intervals (default: 0.95)
digits	Number of decimal places in the output (default: 4).

Value

Results of the multiple imputation analysis for the generalized linear model. For computing covariance matrix estimate, the ordinary Rubin's rule is adapted to the model variance estimates.

- `coef`: Coefficient estimates; transformed to the exponential scale if `eform=TRUE`.
- `SE`: Standard error estimates for `coef`.
- `CL`: Lower limits of confidence intervals.
- `CU`: Upper limits of confidence intervals.
- `df`: Degree of freedom for the t-approximation.
- `P-value`: P-values for the coefficient tests.

References

Little, R. J., and Rubin, D. B. (2019). *Statistical Analysis with Missing Data*, 3rd edition. New York: Wiley.

Examples

```
library("mice")

data(exdata03)

exdata03$x2 <- factor(exdata03$x2)
exdata03$x3 <- factor(exdata03$x3)
exdata03$x4 <- factor(exdata03$x4)

ice5 <- mice(exdata03,m=5)
# For illustration. m should be >=100.

mi_glm(ice5, y ~ x1 + x2 + x3 + x4, family=binomial, eform=TRUE)
# Logistic regression analysis
# Coefficient estimates are translated to odds ratio scales

mi_glm(ice5, x1 ~ x2 + x3 + x4, family=gaussian)
# Ordinary least-squares regression analysis with the model variance estimator
```

---

mi_rqlm *Multiple imputation analysis for modified Poisson and least-squares regressions*

Description

Multiple imputation analysis for modified Poisson and least-squares regressions is performed for the imputed datasets generated by `mice` function in `mice` package. For computing covariance matrix estimate, the ordinary Rubin's rule is adapted to the sandwich variance estimates. Its validity is checked by several simulation studies for general GEE applications by Beunckens et al. (2008), Birhanu et al. (2011) and Yoo (2010).

Usage

```
mi_rqlm(ice, formula, family=poisson, eform=FALSE, cl=0.95, digits=4)
```

Arguments

`ice`	An output object of `mice` function in `mice` package.
`formula`	An object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be fitted.
`family`	A description of the error distribution and link function to be used in the model. `gaussian`: Modified least-squares regression. `poisson`: Modified Poisson regression.
`eform`	A logical value that specify whether the outcome should be transformed by exponential function (default: `FALSE`)
`cl`	Confidence level for calculating confidence intervals (default: 0.95)
`digits`	Number of decimal places in the output (default: 4).

Value

Results of the multiple imputation analysis for modified Poisson and least-squares regressions. For computing covariance matrix estimate, the ordinary Rubin's rule is adapted to the sandwich variance estimates.

- `coef`: Coefficient estimates; transformed to the exponential scale if `eform=TRUE`.
- `SE`: Robust standard error estimates for `coef`.
- `CL`: Lower limits of confidence intervals.
- `CU`: Upper limits of confidence intervals.
- `df`: Degree of freedom for the t-approximation.
- `P-value`: P-values for the coefficient tests.

References

Aloisio, K. M., Swanson, S. A., Micali, N., Field, A., and Horton, N. J. (2014). Analysis of partially observed clustered data using generalized estimating equations and multiple imputation. *Stata Journal*, **14**, 863-883.

Beunckens, C., Sotto, C., and Molenberghs., G. (2008). A simulation study comparing weighted estimating equations with multiple imputation based estimating equations for longitudinal binary data. *Computational Statistics and Data Analysis*, **52**, 1533-1548.

Birhanu, T., Molenberghs, G., Sotto, C., and Kenward, M. G. (2011). Doubly robust and multiple-imputation-based generalized estimating equations. *Journal of Biopharmaceutical Statistics*, **21**, 202-225.

Little, R. J., and Rubin, D. B. (2019). *Statistical Analysis with Missing Data*, 3rd edition. New York: Wiley.

Yoo, B. (2010). The impact of dichotomization in longitudinal data analysis: a simulation study. *Pharmaceutical Statistics*, **9**, 298-312.

Examples

```
library("mice")

data(exdata03)

exdata03$x2 <- factor(exdata03$x2)
exdata03$x3 <- factor(exdata03$x3)
exdata03$x4 <- factor(exdata03$x4)

ice5 <- mice(exdata03,m=5)
# For illustration. m should be >=100.

mi_rqlm(ice5, y ~ x1 + x2 + x3 + x4, family=poisson, eform=TRUE)
# Modifed Poisson regression analysis
# Coefficient estimates are translated to risk ratio scales

mi_rqlm(ice5, y ~ x1 + x2 + x3 + x4, family=gaussian)
# Modifed least-squares regression analysis
```

`qesci.ls` *Calculating confidence interval for modified least-squares regression based on the quasi-score test*

Description

Recent studies revealed the robust standard error estimates of the modified least-squares regression analysis are generally biased under small or moderate sample settings. To adjust the bias and to provide more accurate confidence intervals, confidence interval and P-value of the test for risk difference by modified least-squares regression are calculated based on the quasi-score test of Noma and Gosho (2024).

Usage

```
qesci.ls(formula, data, x.name=NULL, cl=0.95, C0=10^-5, digits=4)
```

Arguments

formula	An object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be fitted.
data	A data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the variables in the model.
x.name	The variable name that the confidence interval is calculated for the regression coefficient; should be involved in `formula` as an explanatory variable. Specify as a character object.
cl	Confidence level for calculating confidence intervals (default: 0.95)
C0	A tuning parameter to control the precisions of numerical computations of confidence limits (default: 10^-5).
digits	Number of decimal places in the output (default: 4).

Value

Results of the modified least-squares analyses are presented. Three objects are provided: Results of the modified least-squares regression with the Wald-type approximation by `rqlm`, quasi-score confidence interval for the corresponding covariate, and P-value for the quasi-score test of `RD=0`.

References

Noma, H. and Gosho, M. (2025). Finite-sample improved confidence intervals based on the estimating equation theory for the modified Poisson and least-squares regressions. *Epidemiologic Methods* **14**, 20240030.

Examples

```
data(exdata01)

qesci.ls(y ~ x1 + x2 + x3 + x4, data=exdata01, "x3")
```

---

qesci.pois — *Calculating confidence interval for modified Poisson regression based on the quasi-score test*

Description

Recent studies revealed the risk ratio estimates and robust standard error estimates of the modified Poisson regression analysis are generally biased under small or moderate sample settings. To adjust the bias and to provide more accurate confidence intervals, confidence interval and P-value of the test for risk ratio by modified Poisson regression are calculated based on the quasi-score test of Noma and Gosho (2024).

Usage

```
qesci.pois(formula, data, x.name=NULL, eform=FALSE, cl=0.95, C0=10^-5, digits=4)
```

Arguments

`formula`	An object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be fitted.
`data`	A data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the variables in the model.
`x.name`	The variable name that the confidence interval is calculated for the regression coefficient; should be involved in `formula` as an explanatory variable. Specify as a character object.
`eform`	A logical value that specify whether the outcome should be transformed by exponential function (default: `FALSE`)
`cl`	Confidence level for calculating confidence intervals (default: 0.95)
`C0`	A tuning parameter to control the precisions of numerical computations of confidence limits (default: 10^-5).
`digits`	Number of decimal places in the output (default: 4).

Value

Results of the modified Poisson analyses are presented. Three objects are provided: Results of the modified Poisson regression with the Wald-type approximation by `rqlm`, quasi-score confidence interval for the corresponding covariate, and P-value for the quasi-score test of `RR=1`.

References

Noma, H. and Gosho, M. (2025). Finite-sample improved confidence intervals based on the estimating equation theory for the modified Poisson and least-squares regressions. *Epidemiologic Methods* **14**, 20240030.

Examples

```
data(exdata01)

qesci.pois(y ~ x1 + x2 + x3 + x4, data=exdata01, "x3", eform=TRUE)
```

`qlogist` *Augmented (modified) logistic regression analyses for estimating risk ratio*

Description

Logistic regression with augmented pseudo-observations for estimating risk ratios is performed. This function is handled by a similar way with `lm` or `glm`. Also, the resultant coefficients and confidence limits can be transformed to exponential scales by specifying `eform`. The Morel-Bokossa-Neerchaal-type small-sample corrected estimator is adopted for standard error estimation as the default method.

Usage

```
qlogist(formula, data, eform=TRUE, cl=0.95, digits=4, var.method="MBN")
```

Arguments

`formula`	An object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be fitted.
`data`	A data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the variables in the model.
`eform`	A logical value that specify whether the outcome should be transformed by exponential function (default: `TRUE`)
`cl`	Confidence level for calculating confidence intervals (default: 0.95)
`digits`	Number of decimal places in the output (default: 4).
`var.method`	Method for estimating standard errors. Standard robust variance estimator (`standard`), Morel-Bokossa-Neerchaal-type corrected estimator (`MBN`), Gosho-Sato-Takeuchi-type corrected estimator (`GST`), and Wang-Long-type corrected estimator (`WL`) are available (default: `MBN`).

Value

Results of the augmented (modified) logistic regression analysis.

- `coef`: Coefficient estimates; transformed to the exponential scale if `eform=TRUE`.
- `SE`: Robust standard error estimates for `coef`.
- `CL`: Lower limits of confidence intervals.
- `CU`: Upper limits of confidence intervals.
- `P-value`: P-values for the coefficient tests.

References

Diaz-Quijano, F. A. (2012). A simple method for estimating relative risk using logistic regression. *BMC Medical Research Methodology* **12**, 14.

Gosho, M., Sato, Y., and Takeuchi, H. (2014). Robust covariance estimator for small-sample adjustment in the generalized estimating equations: a simulation study. *Science Journal of Applied Mathematics and Statistics* **2**, 20-25.

Morel, J. G., Bokossa, M., and Neerchal, N. (2003). Small sample correction for the variance of GEE estimators. *Biometrical Journal* **45**, 395-409.

Noma, H. (2025). Robust variance estimators for risk ratio estimators from logistic regression in cohort and case-cohort studies. Forthcoming.

Noma, H., and Gosho, M. (2025). Logistic mixed-effects model analysis with pseudo-observations for estimating risk ratios in clustered binary data analysis. *Statistics in Medicine* **44**, e70280.

Schouten, E. G., Dekker, J. M., Kok, F. J., et al. (1993). Risk ratio and rate ratio estimation in case-cohort designs: hypertension and cardiovascular mortality. *Statistics in Medicine* **12**, 1733-1745.

Shiiba, H., and Noma, H. (2025). Confidence intervals of risk ratios for the augmented logistic regression with pseudo-observations. *Stats* **8**, 83.

Wang, M., and Long, Q. (2011). Modified robust variance estimator for generalized estimating equations with improved small-sample performance. *Statistics in Medicine* **30**, 1278-1291.

Examples

```
data(exdata02)

qlogist(y ~ x1 + x2 + x3 + x4, data=exdata02)
# Augmented logistic regression analysis
# Coefficient estimates are translated to risk ratio scales
# MBN robust variance estimator is adopted.

qlogist(y ~ x1 + x2 + x3 + x4, data=exdata02, var.method="GST")
# GST robust variance estimator is adopted.

qlogist(y ~ x1 + x2 + x3 + x4, data=exdata02, var.method="WL")
# WL robust variance estimator is adopted.
```

`rqlm` *Modified Poisson and least-squares regression analyses for binary outcomes*

Description

Modified Poisson and least-squares regression analyses for binary outcomes are performed. This function is handled by a similar way with `lm` or `glm`. The model fitting to the binary data can be specified by `family`. Also, the resultant coefficients and confidence limits can be transformed to exponential scales by specifying `eform`. The Morel-Bokossa-Neerchaal-type small-sample corrected estimator is adopted for standard error estimation as the default method.

Usage

```
rqlm(formula, data, family=poisson, eform=FALSE, cl=0.95, digits=4,
var.method="MBN")
```

Arguments

`formula`	An object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be fitted.
`data`	A data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the variables in the model.
`family`	A description of the error distribution and link function to be used in the model. `gaussian`: Modified least-squares regression. `poisson`: Modified Poisson regression.
`eform`	A logical value that specify whether the outcome should be transformed by exponential function (default: `FALSE`)
`cl`	Confidence level for calculating confidence intervals (default: 0.95)

`digits`	Number of decimal places in the output (default: 4).
`var.method`	Method for estimating standard errors. Standard robust variance estimator (`standard`), Morel-Bokossa-Neerchaal-type corrected estimator (`MBN`), Gosho-Sato-Takeuchi-type corrected estimator (`GST`), and Wang-Long-type corrected estimator (`WL`) are available (default: `MBN`).

Value

Results of the modified Poisson and least-squares regression analyses.

- `coef`: Coefficient estimates; transformed to the exponential scale if `eform=TRUE`.
- `SE`: Robust standard error estimates for `coef`.
- `CL`: Lower limits of confidence intervals.
- `CU`: Upper limits of confidence intervals.
- `P-value`: P-values for the coefficient tests.

References

Cheung, Y. B. (2007). A modified least-squares regression approach to the estimation of risk difference. *American Journal of Epidemiology* **166**, 1337-1344.

Gosho, M., Ishii, R., Noma, H., and Maruo, K. (2023). A comparison of bias-adjusted generalized estimating equations for sparse binary data in small-sample longitudinal studies. *Statistics in Medicine* **42**, 2711-2727.

Gosho, M., Sato, Y., and Takeuchi, H. (2014). Robust covariance estimator for small-sample adjustment in the generalized estimating equations: a simulation study. *Science Journal of Applied Mathematics and Statistics* **2**, 20-25.

Morel, J. G., Bokossa, M., and Neerchal, N. (2003). Small sample correction for the variance of GEE estimators. *Biometrical Journal* **45**, 395-409.

Noma, H. and Gosho, M. (2025). Finite-sample improved confidence intervals based on the estimating equation theory for the modified Poisson and least-squares regressions. *Epidemiologic Methods* **14**, 20240030.

Noma, H., Sunada, H., and Gosho, M. (2025). Quasi-likelihood ratio tests and the Bartlett-type correction for improved inferences of the modified Poisson and least-squares regressions for binary outcomes. *Statistica Neerlandica* **79**, e70012.

Wang, M., and Long, Q. (2011). Modified robust variance estimator for generalized estimating equations with improved small-sample performance. *Statistics in Medicine* **30**, 1278-1291.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1-25.

Zou, G. (2004). A modified poisson regression approach to prospective studies with binary data. *American Journal of Epidemiology* **159**, 702-706.

Examples

```
data(exdata02)

rqlm(y ~ x1 + x2 + x3 + x4, data=exdata02, family=poisson, eform=TRUE)
# Modifed Poisson regression analysis
# Coefficient estimates are translated to risk ratio scales
# MBN robust variance estimator is adopted.

rqlm(y ~ x1 + x2 + x3 + x4, data=exdata02, family=gaussian)
# Modifed least-squares regression analysis

rqlm(y ~ x1 + x2 + x3 + x4, data=exdata02, family=gaussian, digits=3)
# Modifed least-squares regression analysis
# Number of decimal places can be changed by specifying "digits"

rqlm(y ~ x1 + x2 + x3 + x4, data=exdata02, family=poisson, eform=TRUE, var.method="GST")
# Modifed Poisson regression analysis
# Coefficient estimates are translated to risk ratio scales
# GST robust variance estimator is adopted.

rqlm(y ~ x1 + x2 + x3 + x4, data=exdata02, family=poisson, eform=TRUE, var.method="WL")
# Modifed Poisson regression analysis
# Coefficient estimates are translated to risk ratio scales
# WL robust variance estimator is adopted.
```

---

stabwt *Calculating stabilized weights for IPW analysis: Single time point*

Description

This function calculates stabilized weights for IPW analysis using logistic regression model. Both of untruncated and utruncated weights are provided.

Usage

```
stabwt(formula, data, trunc=c(0.01,0.99), digits=4)
```

Arguments

formula	An object of class "formula" (or one that can be coerced to that class): a symbolic description of the logistic regression model to be fitted.
data	A data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the variables in the model.
trunc	Quantiles to be truncated the weights (default: 0.01, 0.99).
digits	Number of decimal places in the output (default: 4).

Value

Truncated and untruncated stabilized weights are calculated.

- `sw1`: Untruncated stabilized weights.
- `sw2`: Truncated stabilized weights.

References

Cole, S. R., and Hernan, M. A. (2008). Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology* **168**, 656-664.

Westreich, D., Edwards, J. K., Lesko, C. R., Stuart, E., and Cole, S. R. (2017). Transportability of trial results using inverse odds of sampling weights. *American Journal of Epidemiology* **186**, 1010-1014.

Examples

```
data(exdata02)

stabwt(x2 ~ x1 + x3 + x4, data=exdata02)
```

`stabwtlong` *Calculating stabilized weights for IPW analysis: Longitudinal data*

Description

This function calculates stabilized weights for IPW analysis of longitudinal data using logistic regressions. Both of untruncated and utruncated weights are provided.

Usage

```
stabwtlong(formula_denom, formula_num, data, trunc=c(0.01,0.99), digits=4)
```

Arguments

`formula_denom`	An object of class "formula" (or one that can be coerced to that class): a symbolic description of the denominator logistic regression model to be fitted.
`formula_num`	An object of class "formula" (or one that can be coerced to that class): a symbolic description of the numerator logistic regression model to be fitted.
`data`	A data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the variables in the model.
`trunc`	Quantiles to be truncated the weights (default: 0.01, 0.99).
`digits`	Number of decimal places in the output (default: 4).

Value

Truncated and untruncated stabilized weights are calculated.

- `sw1`: Untruncated stabilized weights.
- `sw2`: Truncated stabilized weights.

References

Cole, S. R., and Hernan, M. A. (2008). Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology* **168**, 656-664.

Westreich, D., Edwards, J. K., Lesko, C. R., Stuart, E., and Cole, S. R. (2017). Transportability of trial results using inverse odds of sampling weights. *American Journal of Epidemiology* **186**, 1010-1014.

Examples

```
data(exdata04)

stabwtlong(formula_denom = A ~ L1 + L2 + L3, formula_num = A ~ L1, data = exdata04)
```

`stabwtmulti` *Calculating stabilized weights for IPW analysis: Single time point (for more than 3 groups)*

Description

This function calculates stabilized weights for IPW analysis using logistic regression model. Both of untruncated and utruncated weights are provided.

Usage

```
stabwtmulti(formula, data, trunc=c(0.01,0.99), digits=4)
```

Arguments

`formula`	An object of class "formula" (or one that can be coerced to that class): a symbolic description of the logistic regression model to be fitted.
`data`	A data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the variables in the model.
`trunc`	Quantiles to be truncated the weights (default: 0.01, 0.99).
`digits`	Number of decimal places in the output (default: 4).

Value

Truncated and untruncated stabilized weights are calculated.

- `sw1`: Untruncated stabilized weights.
- `sw2`: Truncated stabilized weights.

References

Cole, S. R., and Hernan, M. A. (2008). Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology* **168**, 656-664.

Westreich, D., Edwards, J. K., Lesko, C. R., Stuart, E., and Cole, S. R. (2017). Transportability of trial results using inverse odds of sampling weights. *American Journal of Epidemiology* **186**, 1010-1014.

SumStat — *Creating summary table for IPTW analysis using stabilized weights*

Description

Summary statistics are computed before and after weighting for IPTW analyses. For all covariates included in the weight model, the mean, standard deviation, and standardized mean difference (SMD) are calculated. Weighting can be performed using either stabilized or unstabilized weights, with optional truncation. For binary covariates, summary statistics are computed using unbiased estimators of the binomial mean and variance.

Usage

```
SumStat(formula, data, trunc=c(0.01,0.99), digits=3)
```

Arguments

`formula`	An object of class "formula" (or one that can be coerced to that class): a symbolic description of the logistic regression model to be fitted.
`data`	A data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the variables in the model.
`trunc`	Quantiles to be truncated the weights (default: 0.01, 0.99).
`digits`	Number of decimal places in the output (default: 4).

Value

Summary statistics for unadjusted and IPTW-adjusted populations are provided. For each covariate, the mean and standard deviation, as well as the standardized mean difference (SMD), were calculated. Truncated stabilized weights were used for weighting. Setting the `trunc` argument to c(0, 1) allows weighting with untruncated stabilized weights.

- `mean0`: Mean for group 0.
- `mean1`: Mean for group 1.
- `sd0`: SD for group 0.
- `sd1`: SD for group 1.
- `SMD`: SMD for the original population.
- `wmean0`: IPTW-weighted mean for group 0.

- wmean1: IPTW-weighted mean for group 1.
- wsd0: IPTW-weighted SD for group 0.
- wsd1: IPTW-weighted SD for group 1.
- wSMD: IPTW-weighted SMD for the adjusted pseudo-population.
- type: Type of the variable.

References

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine* **28**, 3083-3107.

Examples

```
data(exdata02)

SumStat(x2 ~ x1 + x3 + x4, data=exdata02)
```

---

ttemsm *Pooled logistic regression for target trial emulation*

Description

This function implements pooled logistic regression for use in target trial emulation. Before running the function, the user must prepare an analysis dataset by stacking the sequential-trial datasets in long format. If inverse probability weights (such as IPCW) are required, the corresponding weight variable should be included in the dataset in advance. The regression model is specified through the formula argument, and the individual identifier must be provided via the id argument, so that cluster-robust standard errors are computed across repeated trials for each individual. When a weight variable is specified through weights, the function performs inverse probability weighting. The output includes point estimates of the hazard ratio, corresponding confidence intervals, and P-values.

Usage

```
ttemsm(formula, data, id, weight, family=quasibinomial(link="cloglog"),
eform=TRUE, cl=0.95, digits=4, var.method="MBN")
```

Arguments

formula	An object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be fitted.
data	A data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the variables in the model.
id	id variable for individual participants containing in the data object.
weight	weight variable for individual participants containing in the data object.

`family` A description of the error distribution and link function to be used in the model. When the goal is to estimate hazard ratios using a discrete-time Cox model, the complementary log–log (cloglog) link should be used. When the objective is to estimate risk differences or cumulative incidence (e.g., via the g-formula or standardization), the logit link should be used.

`eform` A logical value that specify whether the outcome should be transformed by exponential function (default: `TRUE`)

`cl` Confidence level for calculating confidence intervals (default: 0.95)

`digits` Number of decimal places in the output (default: 4).

`var.method` Method for estimating standard errors. Standard cluster-robust variance estimator (`standard`) and Morel-Bokossa-Neerchaal-type corrected cluster-robust estimator (`MBN`) are available (default: `MBN`).

Value

Results of the pooled logistic regression analysis.

- `coef`: Coefficient estimates; transformed to the exponential scale if `eform=TRUE`.
- `SE`: Cluster-robust standard error estimates for `coef`.
- `CL`: Lower limits of confidence intervals.
- `CU`: Upper limits of confidence intervals.
- `P-value`: P-values for the coefficient tests.

References

Gosho, M., Ishii, R., Noma, H., and Maruo, K. (2023). A comparison of bias-adjusted generalized estimating equations for sparse binary data in small-sample longitudinal studies. *Statistics in Medicine* **42**, 2711-2727.

Hernan, M. A., Alonso, A., Logan, R., et al. (2008). Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology* **19**, 766-779.

Hernan, M. A., Dahabreh, I. J., Dickerman, B. A., and Swanson, S. A. (2025). The Target Trial Framework for Causal Inference From Observational Data: Why and When Is It Helpful? *Annals of Internal Medicine* **178**, 402-407.

Hernan, M. A., Wang, W., and Leaf, D. E. (2022). Target Trial Emulation: A Framework for Causal Inference From Observational Data. *JAMA* **328**, 2446-2447.

Morel, J. G., Bokossa, M., and Neerchal, N. (2003). Small sample correction for the variance of GEE estimators. *Biometrical Journal* **45**, 395-409.

Examples

```
data(exdata04)

ttemsm( Y ~ A + L1 + L2 + L3 + time + I(time^2) + trial,
  data    = exdata04, id = ID, weight = w_pp,
  family = quasibinomial(link="cloglog"),
```

```
  eform   = TRUE, cl = 0.95, var.method="standard")
# Pooled logistic regression for target trial emulation with cloglog link
# For estimating hazard ratios using a discrete-time Cox model

ttemsm( Y ~ A + L1 + L2 + L3 + time + I(time^2) + trial,
  data    = exdata04, id = ID, weight = w_pp,
  family = quasibinomial(link="logit"),
  eform   = TRUE, cl = 0.95, var.method="standard")
# Pooled logistic regression for target trial emulation with logit link
# For estimating RDs or cumulative incidence (e.g., via the g-formula)

ttemsm( Y ~ A + L1 + L2 + L3 + time + I(time^2) + trial,
  data    = exdata04, id = ID, weight = w_pp,
  eform   = TRUE, cl = 0.95, var.method="MBN")
# Pooled logistic regression for target trial emulation with cloglog link
# Morel-Bokossa-Neerchaal-type corrected SE estimator is used.
```

Index