Package 'rqlm'

July 23, 2025

Type Package

Title Modified Poisson and Least-Squares Regressions for Binary
Outcome and Their Generalizations

Version 2.3-1

Date 2025-01-12

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

Description Modified Poisson and least-squares regression analyses for binary outcomes of Zou (2004) <doi:10.1093/aje/kwh090> 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 (Noma and Gosho (2024) <Forthcoming>).

Depends R (>= 3.5.0)

Imports stats, MASS, sandwich, mice

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 2025-01-12 08:00:01 UTC

Contents

---

rqlm-package   *The 'rqlm' package.*

Description

Modified Poisson 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 (Noma and Gosho, 2024).

References

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

Noma, H. and Gosho, M. (2024). Bootstrap confidence intervals based on quasi-likelihood estimating functions for the modified Poisson and least-squares regressions for binary outcomes. Forthcoming.

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. (2024). Bootstrap confidence intervals based on quasi-likelihood estimating functions for the modified Poisson and least-squares regressions for binary outcomes. Forthcoming.

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. (2024). Bootstrap confidence intervals based on quasi-likelihood estimating functions for the modified Poisson and least-squares regressions for binary outcomes. Forthcoming.

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 linear regression model and the generalized linear model can be calculated using the ordinary model variance estimators. Through simply entering the output objects of `lm` or `glm`, 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` or `glm`.
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
```

---

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.

---

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. (2024). Bootstrap confidence intervals based on quasi-likelihood estimating functions for the modified Poisson and least-squares regressions for binary outcomes. Forthcoming.

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. (2024). Bootstrap confidence intervals based on quasi-likelihood estimating functions for the modified Poisson and least-squares regressions for binary outcomes. Forthcoming.

Examples

```
data(exdata01)

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

---

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 standard error estimates are calculated using the standard robust variance estimator by sandwich package.

Usage

```
rqlm(formula, data, family=poisson, eform=FALSE, cl=0.95, 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.
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 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.

Noma, H. and Gosho, M. (2024). Bootstrap confidence intervals based on quasi-likelihood estimating functions for the modified Poisson and least-squares regressions for binary outcomes. Forthcoming.

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
```

```
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"
```

Index