

rqlm: R package for implementing the modified Poisson and least-squares regression analyses with simple commands

Hisashi Noma

The Institute of Statistical Mathematics, Tokyo, Japan

e-mail: noma@ism.ac.jp

ORCID: <https://orcid.org/0000-0002-2520-9949>

January 21, 2024

1. Introduction

Logistic regression has been widely used for the multivariate analyses of binary outcomes in clinical and epidemiological studies. However, the resultant odds ratio estimator cannot be directly interpreted as an effect measure and is only be interpreted as an approximation of risk ratio when the event frequency is low (Greenland, 1987; Nurminen, 1995).

To address this problem, Zou (2004) proposed a modified Poisson regression that provide risk ratio estimates directly. Since this method can be performed by ordinary Poisson regression tools in standard statistical software (e.g., R, SAS, Stata), and has increasingly been adopted in practice. The original paper of Zou (2004) is featured one of the most cited papers in the 100 year history of *The American Journal of Epidemiology* and has currently been cited more than 8,000 times by Google Scholar. Also, Cheung (2007) proposed a modified least-squares regression that provide risk difference estimates directly.

However, these methods require some complicated calculations of robust standard error estimates and their applications for computations of the confidence intervals and P-values. The `rqlm` package involve an easy-to-handle function `rqlm` that enables all of these computations by simple commands.

2. Installation

The `rqlm` package can be installed from CRAN by the following command,

```
> install.packages("rqlm")
```

3. Loading the package

After the installation, users can load the package via the following command.

```
> library("rqlm")
```

4. Example dataset

Here, we use an example dataset involved in rqlm package for illustration.

```
data(exdata02)
head(exdata02)
  y      x1  x2  x3  x4
1 0 29.35836  1  1  1
2 0 27.41645  0  1  0
3 0 37.65188  0  1  1
4 0 31.06593  0  1  0
5 0 26.46533  1  1  0
6 0 29.13499  0  1  0
```

exdata02 is a cohort data that involves 1,200 participants data. y is the outcome variable (= 0, 1) and x1, x2, x3, x4 are the covariates.

5. Modified Poisson regression analysis

The modified Poisson regression can be performed using rqlm function in the rqlm package. An example command is as follows:

```
rqlm(y ~ x1 + x2 + x3 + x4, data=exdata02, family=poisson,
eform=TRUE)
# Modified Poisson regression analysis
# Coefficient estimates are translated to risk ratio scales
```

	exp(coef)	SE	CL	CU	P-value
(Intercept)	0.3051	0.2588	0.1837	0.5066	0.0000
x1	0.9921	0.0079	0.9769	1.0075	0.3116
x2	1.2153	0.1150	0.9701	1.5225	0.0899
x3	0.6632	0.1260	0.5181	0.8489	0.0011
x4	1.2356	0.1911	0.8495	1.7971	0.2684

The syntax is similar to lm or glm function of basic functions of R. At first, users should specify the formula such that $y \sim x1 + x2 + x3 + x4$. Then, the dataset object should be specified by data. Also, the distribution type of outcome variable should be

specified by the argument `family`; the usage is the same with that of `glm`. For modified Poisson regression, this argument should be `family=poisson`. Also, for the modified Poisson regression, the regression coefficient estimates should finally be transformed to exponential scale to obtain risk ratio estimates. If an argument `eform=TRUE` is added, the results are transformed to exponential scale.

For the result table, the coefficient estimates (if `eform=TRUE`, it is transformed to an exponential scale), SE (standard error estimates), CL/CU (lower and upper limits of confidence intervals; the confidence level is 95% (default), but can be changed; if `eform=TRUE`, they are transformed to exponential scales), P-value are presented.

6. Modified least-squares regression analysis

Similarly, the modified least-squares regression analysis can be performed using `rqlm` function,

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

	coef	SE	CL	CU	P-value
(Intercept)	0.2947	0.0557	0.1854	0.4039	0.0000
x1	-0.0016	0.0016	-0.0047	0.0015	0.3191
x2	0.0392	0.0233	-0.0065	0.0849	0.0927
x3	-0.0902	0.0303	-0.1496	-0.0307	0.0029
x4	0.0420	0.0406	-0.0376	0.1215	0.3013

The usage is the same as the modified Poisson regression. Please change the argument `family` to `family=gaussian`. Also, for the modified least-squares regression, the regression coefficients can be directly interpreted as the risk difference, please set `eform=FALSE` (default).

6. Change of the number of decimal places

For the `rqlm` function, the number of decimal places can be changed by the argument `digits` (default: 4). If it is changed to 3, the output is changed as follows,

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

	coef	SE	CL	CU	P-value
(Intercept)	0.295	0.056	0.185	0.404	0.000
x1	-0.002	0.002	-0.005	0.002	0.319
x2	0.039	0.023	-0.006	0.085	0.093
x3	-0.090	0.030	-0.150	-0.031	0.003
x4	0.042	0.041	-0.038	0.122	0.301

An R file that provides the example codes above is available at the following URL:

https://www.ism.ac.jp/~noma/ex_rqlm.r

When you publish the data obtained from analyses using `rqlm` package, I am glad if you would refer the following software:

Noma, H. (2024). “`rqlm`: Modified Poisson and least-squares regressions for binary outcome.” R package version 1.2-1. <https://cran.r-project.org/web/packages/rqlm/>

References

- Cheung, Y. B. (2007). A modified least-squares regression approach to the estimation of risk difference. *American Journal of Epidemiology* **166**, 1337-1344.
- Greenland, S. (1987). Interpretation and choice of effect measures in epidemiologic analysis. *American Journal of Epidemiology* **125**, 761-768.
- Nurminen, M. (1995). To use or not to use the odds ratio in epidemiologic analyses. *European Journal of Epidemiology* **11**, 365-371.
- Zou, G. (2004). A modified poisson regression approach to prospective studies with binary data. *American Journal of Epidemiology* **159**, 702-706.