

rqlm: 修正ポアソン回帰と修正最小二乗回帰による解析を行うための
R パッケージ

野間 久史

統計数理研究所

e-mail: noma@ism.ac.jp

URL: <https://www.ism.ac.jp/~noma/>

2024 年 1 月 21 日

1. 修正ポアソン回帰・修正最小二乗回帰とは？

臨床研究・疫学研究における 2 値アウトカムの解析のための多変量解析では、ロジスティック回帰モデルが長らく慣用されてきました。しかし、ロジスティック回帰モデルから推定されるオッズ比は、集団におけるアウトカムのイベントの頻度が低いときに（概ね 10%未満）、リスク比の近似になる以外、曝露・治療効果の指標としての解釈ができないことが知られています (Greenland, 1987; Nurminen, 1995)。

この問題を解決するために、2004 年の *American Journal of Epidemiology* 誌で、オッズ比ではなく、リスク比そのものを推定するための方法として、修正ポアソン回帰 (modified Poisson regression; Zou, 2004) という多変量解析の方法が提案されました。*American Journal of Epidemiology* 誌の 100 年の歴史の中で、最も多く引用された論文のひとつとされており、現在、Google Scholar での引用回数も 8,000 回を超えています。同じく、リスク差を直接推定することができる方法として、修正最小二乗回帰 (modified least-squares regression; Cheung, 2007) という多変量解析の方法も提案されています。

いずれも、信頼区間の計算に少し複雑なツールを使う必要があり、統計家でない方々には、その実装は必ずしも容易ではありませんが、今回、フリーの統計ソフトウェア R を使って、1 行の簡単なコマンドで、推定値・信頼区間・P 値の計算を行うことを可能としたライブラリ、`rqlm` を作成しました。2024 年 1 月現在、CRAN に公式に登録されており、どなたでも自由に利用することができます。

2. rqlm によるデータ解析の方法

`rqlm` package は、R を立ち上げて、以下のコマンドを入力するだけで、CRAN からインストールすることができます。

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

インストールが完了したら、以下のコマンドで、パッケージを読み込みます。

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

ここでは、`rqlm` package に実装されている `exdata1200` という事例データの解析を通して、修正ポアソン回帰・修正最小二乗回帰での解析方法を紹介します。

```
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` は、1200 人のコホートのデータが含まれている。`y` が、イベントの有無を表す 2 値変数 (0 か 1 の値をとる) であり、`x1`, `x2`, `x3`, `x4` が、説明変数としてモデル化される曝露変数・調整変数です。

まず、修正ポアソン回帰による分析ですが、以下のようなコマンドで解析を行うことができます。

```
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

引数には、`lm` や `glm` と同じく、結果変数と説明変数の組を、式の形にした `formula` を入れます。その後、`data` という引数で、解析対象となるデータセットのオブジェクトを指定します。`family` は、`glm` と使い方はまったく同じです。修正ポアソン回帰を行いたいのであれば、`family=poisson` とします。`eform` は、最終的に得られた回帰係数の推定値を、指数関数で変換するか否かを指定する変数です。修正ポアソン回帰の場合は、指数関数で変換することによって、リスク比の推定値が得られます (`eform=TRUE` とすることで、リスク比の推定値が得られます)。

結果の表には、回帰係数の推定値 (`eform=TRUE` としたら、指数変換したもの)、`SE` (standard error)、`CL/CU` (信頼区間の下上限; デフォルトでは 95%信頼区間 (信頼水準は変更可); `eform=TRUE` としたら、指数変換したもの)、`P-value` (P 値) が出力されます。

同様に、修正最小二乗回帰による分析は、以下のコマンドで実行することができます。

```
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

引数は、修正ポアソン回帰と同様です。`family=gaussian` とすることで、修正最小二乗回帰での解析を行うことができます。修正最小二乗回帰では、回帰係数は、そのまま、リスク差の推定値と解釈することができるので、`eform=FALSE` (default) として分析を行います。結果の表も、修正ポアソン回帰と同様です。

ちなみに、`digits` という引数を使って、小数点以下の桁数を変更することができます。デフォルトでは、`digits=4` となっており、4 桁目までの数値が出力されますが、以下のように変更すると、出力される桁数が変わります。

```
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

以上の事例コマンドを実行するための R ファイルを、以下の URL に公開しています。

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

`rqlm` package を使用した解析結果を公表される際には、以下のソフトウェアを引用していただくと幸いです。

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/>

文 献

- 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.