

修正ポアソン回帰・修正最小二乗回帰による 2値アウトカムデータの多変量解析

野間 久史
統計数理研究所
2024年5月14日
九州大学大学院医学研究院
e-mail: noma@ism.ac.jp
URL: <http://www.ism.ac.jp/~noma/>

1

最近の論文報告ガイドラインにおいて

- ▶ New England Journal of Medicine誌の2019年版の統計ガイドライン：“Odds ratios should be avoided, as they may overestimate the relative risks in many settings and be misinterpreted.”
- ▶ CONSORT2010のチェックリスト：“For binary outcomes, presentation of both absolute and relative effect sizes is recommended.”

Schulz et al. (2010) 2

代表的な統計解析手法

変数の型 解析の目的	2値 Binary	連続 Continuous	生存時間 Survival
分布の記述	頻度集計 分割表	ヒストグラム 平均・SD・相関係数	Kaplan-Meier曲線
単純なグループ比較	カイニ乗検定 Fisherの正確検定 リスク差・リスク比の 推定	t検定, Wilcoxon検定 平均値の差	Logrank検定 人-時間法
多変量解析/回帰モデル (交絡調整)	ロジスティック回帰	線形回帰分析 単回帰・重回帰モデル	Cox回帰

3

本日のお話

- ▶ 結果変数が2値変数（0 or 1）のデータの多変量解析で使われる回帰モデル
- ▶ ロジスティック回帰から推定される治療効果・曝露効果の指標はオッズ比であり、稀な疾病のもと、リスク比の近似になる以外の解釈ができない
- ▶ より直接的な解釈ができるリスク比・リスク差を直接的に推定することができる修正ポアソン回帰・修正最小二乗回帰という方法が、近年、臨床研究の実践において、急速に普及している

4

American Journal of EPIDEMIOLOGY

SOCIETY OF EPIDEMIOLOGIC RESEARCH

JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

Issues More Content ▾ Submit ▾ Purchase Alerts About ▾
American Journal of Epidemiology ▾
Advanced Search

No cover image available

JOURNAL ARTICLE

A Modified Poisson Regression Approach to Prospective Studies with Binary Data

Guangyong Zou

American Journal of Epidemiology, Volume 159, Issue 7, 1 April 2004, Pages 702–706,
<https://doi.org/10.1093/aje/kwh090>

Published: 01 April 2004

PDF
 Split View
 Cite
 Permissions
 Share ▾

Abstract

Relative risk is usually the parameter of interest in epidemiologic and medical studies. In this paper, the author proposes a modified Poisson regression approach (*i.e.*, Poisson regression with a robust error variance) to estimate this effect measure directly. A simple 2-by-2 table is used to justify the validity of this approach. Results from a limited simulation study indicate that this approach is very reliable even with total sample sizes as small as 100. The method is illustrated with two data sets.

[< Previous](#)
[Next >](#)

Article Contents

Abstract

MODIFIED POISSON REGRESSION

ILLUSTRATIVE EXAMPLES

DISCUSSION

ACKNOWLEDGMENTS

References

CITATIONS

VIEWS

ALTMETRIC

More metrics information

Email alerts

[Article activity alert](#)

[Advance article alerts](#)

[New issue alert](#)

Receive exclusive offers and updates
from Oxford Academic

Recommended

On heterogeneity tests based on efficient score
ROBERT E. TARDINE, Biometrika, 1989

Phantom interactions: Use odds ratios or risk
misinterpreting occupancy models
Gavin M Jones et al., The Condor, 2018

Dynamic panel estimation and homogeneity
testing under cross section dependence
Phillips et al., The Econometrics Journal, 2003

EPIDEMIOLOGY


**Society for
Epidemiologic
Research**

**JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH**

[Issues](#)
 [More Content ▾](#)
 [Submit ▾](#)
 [Purchase](#)
 [Alerts](#)
 [About ▾](#)

[American Journal of Epidio ▾](#)
 [Advanced Search](#)

COLLECTION:

100 years

➦ Share

The *American Journal of Epidemiology* (*AJE*) was founded in 1920, with the first issue published in January 1921. In the past 100 years, the *Journal* has gone through many changes, including changes in frequency of publication and even name (from *American Journal of Hygiene* to *American Journal of Epidemiology* in 1965). Through it all, one thing that has remained the same is our dedication to publishing timely, relevant, and impactful research on epidemiology and public health. In the coming year (and beyond), the *AJE* will solidify its mission statement for the future and make updates to provide a better user experience and launch us into the next era.

As William Welch noted in his introduction to the first issue of the *Journal*, “The history of public health administration and of preventive medicine exemplifies perhaps more strikingly than any other field the dependence of successful practice upon accurate knowledge.” We at the *AJE* are proud to have assisted in the publication and dissemination of that knowledge, and we are looking forward to the next 100 years. Please join us in our celebration!

Read the 100 years editorial in the January 2021 issue

Our most cited papers

Discover the most cited articles of all times from the *American Journal of Epidemiology*.

MOST CITED

A Modified Poisson Regression Approach to Prospective Studies with Binary Data

https://academic.oup.com/aje/pages/100-years

ロジスティック回帰分析

- ▶ 結果変数が2値変数（0 or 1）のデータの変量解析で使われてきた回帰モデル
- ▶ $\log\left(\frac{\Pr(Y=1)}{1-\Pr(Y=1)}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$
- ▶ 回帰係数の $\beta_1, \dots, \beta_p$ は、対応する説明変数 $x_1, \dots, x_p$ が 1 増加するごとのイベント発生確率の対数オッズ比（オッズ比を対数変換したもの）に対応する
- ▶ 回帰係数の推定値・信頼区間を指数関数で変換し、オッズ比の推定値・信頼区間を報告することが一般的である

7

オッズ比とは？

- ▶ オッズとは？
 - ▶ ギャンブルで、勝つ確率と負ける確率の「比」をとったもので、何倍くらい勝てる見込みが高いかを示す指標
- ▶ イベント発生確率のオッズ比（odds-ratio）
 - ▶ $OR = \frac{\Pr(Y=1|X=1)/\{1-\Pr(Y=1|X=1)\}}{\Pr(Y=1|X=0)/\{1-\Pr(Y=1|X=0)\}}$
 - ▶ 「治療を受けたときのイベント発生確率のオッズ」と「治療を受けなかったときのイベント発生確率のオッズ」の比

8

リスク比とは？

- ▶ リスク (Risk)
 - ▶ 疫学用語としては、「(好ましくない) イベントの起こる確率」
- ▶ リスク比 (risk-ratio)
 - ▶ $RR = \frac{\Pr(Y=1|X=1)}{\Pr(Y=1|X=0)}$
 - ▶ 「治療を受けたときのイベント発生確率」と「治療を受けなかったときのイベント発生確率」の比
- ▶ イベント発生の相対リスクの指標として、直接的に解釈することができる

9

リスク差とは？

- ▶ リスク差 (risk-difference)
 - ▶ $RD = \Pr(Y = 1|X = 1) - \Pr(Y = 1|X = 0)$
 - ▶ 「治療を受けたときのイベント発生確率」と「治療を受けなかったときのイベント発生確率」の差
- ▶ イベント発生の絶対リスクの指標として、直接的に解釈することができる

10

オッズ比の意味・解釈とは？

- ▶ 「オッズの比」は、その定義自体、疫学的に意味のある指標として解釈を与えることは難しい
- ▶ 1980年代頃に、Sander Greenland教授を中心に、あらゆる観点から、オッズ比の意味を見出すための研究が行われた（それこそ、反事実モデルに基づく因果推論などを使って）
- ▶ しかし、結論として、治療・曝露効果の大きさを測る指標として、直接的な解釈はできない指標であるとされてきた

Greenland (1987), Nurminen (1995) 11

唯一、解釈ができるのは

- ▶ イベントの発生リスクが小さく、稀な疾病を対象とした研究の場合には、

$$\begin{aligned} \text{▶ OR} &= \frac{\Pr(Y=1|X=1)/\{1-\Pr(Y=1|X=1)\}}{\Pr(Y=1|X=0)/\{1-\Pr(Y=1|X=0)\}} && \text{分母のほうがほぼ1に！} \\ &\approx \frac{\Pr(Y=1|X=1)}{\Pr(Y=1|X=0)} \end{aligned}$$

- ▶ オッズ比とリスク比が似たような値になる
- ▶ リスク比の近似として、オッズ比を用いることができる！
- ▶ Greenland (1987) では、オッズ比には、それ以外の効果の指標としての解釈はないと結論づけている

12

んん？だとすると、おかしいのでは？

- ▶ それならば、オッズ比ではなく、リスク比をすべての研究で使えばよいのではないか？
- ▶ まさにその通り！
- ▶ 問題点①：しかし、リスク比・リスク差の回帰モデルを実装しようとする、{0, 1} のデータをモデル化する確率分布（2項分布）の数学的な性質上、ロジスティック回帰モデル以外の回帰モデルの多くは、最尤推定値が定義できないという問題がある！

McNutt et al. (2003), Wallenstein and Bodian (1987) 13

リスク比・リスク差を推定する回帰モデル

- ▶ ロジスティック回帰モデル
 - ▶ $\log\left(\frac{\Pr(Y=1)}{1-\Pr(Y=1)}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$
- ▶ 2項回帰モデル（対数リンク・恒等リンク）
 - ▶ $\log\{\Pr(Y = 1)\} = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$
 - ▶ $\Pr(Y = 1) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$
 - ▶ それぞれ、回帰係数は、リスク比（対数変換したもの）、リスク差に対応する
 - ▶ これらを使えばよいのでは？？

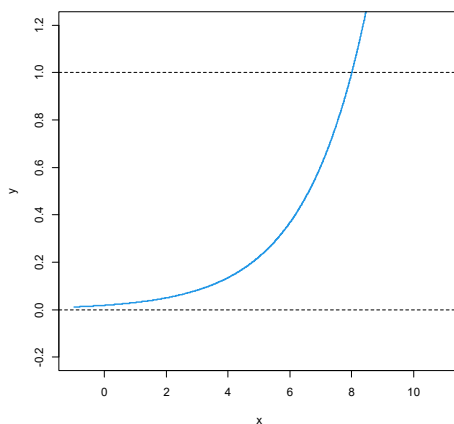
14

Simple counter examples

- ▶ 1つだけ連続な説明変数があるケースにおいて
- ▶ リスク比の回帰モデル
 - ▶ $\Pr(Y = 1) = \exp\{\beta_0 + \beta_1 x_1\}$
- ▶ リスク差の回帰モデル
 - ▶ $\Pr(Y = 1) = \beta_0 + \beta_1 x_1$
- ▶ ロジスティック回帰モデル
 - ▶ $\Pr(Y = 1) = \frac{\exp\{\beta_0 + \beta_1 x_1\}}{1 + \exp\{\beta_0 + \beta_1 x_1\}}$

15

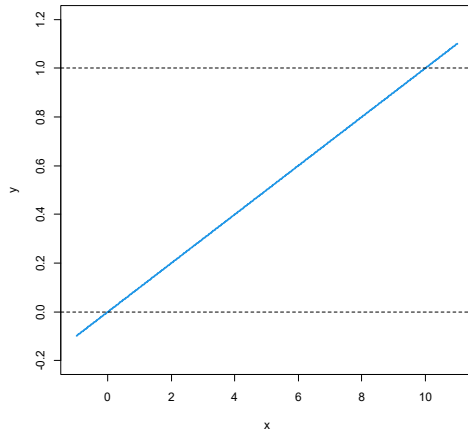
リスク比の回帰モデル



$\exp(\text{XXXX})$ という関数になるので、下限は 0 より大きな値になるが、上限は 1 よりも大きくなる。
[0, 1] の範囲に、関数の値は一般的に収まらない。
もともとは、「確率」をモデル化するものなので、
[0, 1] に収まらないと、論理的定義が成り立たない

16

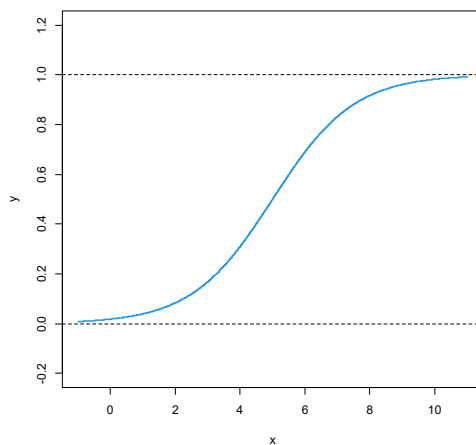
リスク差の回帰モデル



普通の1次関数になるので、下限・上限ともに、 $[0, 1]$ の範囲を超えてしまう。
もともとは、「確率」をモデル化するものなので、 $[0, 1]$ に収まらないと、論理的定義が成り立たない

17

ロジスティック回帰モデル



ロジスティック関数という関数を使うため、下限・上限は、 $[0, 1]$ の範囲に必ず収まる。
「確率」をモデル化する関数モデルとして、望ましい性質を備えている。
これにより、ロジスティック回帰の最尤推定値は、原則として、必ず求められる。

18

現代の見地からはそんなに重要でないが

- ▶ 問題点②：歴史的には、コホート研究とケースコントロール研究が、疫学研究の2つの主要な研究デザインであった
- ▶ ケースコントロール研究では、原則として、リスク差・リスク比を直接的に推定することはできない
- ▶ しかし、オッズ比のみは推定することができる（ロジスティック回帰も使える；もちろん、稀な疾病の時のみ）
- ▶ どちらのデザインでも、共通の指標として使える、オッズ比が、学問体系の発展途上では「簡単であるため」好まれてきたという経緯もあるようである

19

修正ポアソン回帰分析

- ▶ リスク比を推定するための2項回帰（対数リンク）の代替的な方法として、Zou (2004) によって開発された方法
- ▶ 単純に、2値アウトカムのデータに、ポアソン回帰モデルをあてはめて回帰係数を推定するというだけの方法
 - ▶ $\log\{E(Y)\} = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$
- ▶ ポアソン分布は、もともとカウントデータや率データの解析に使われる分布（0, 1, 2, ...という非負の整数値をとる分布）であり、{0, 1}の2値にしか値をとらないデータとはまったく異なる分布をとるので、的外れなモデルになるのでは？

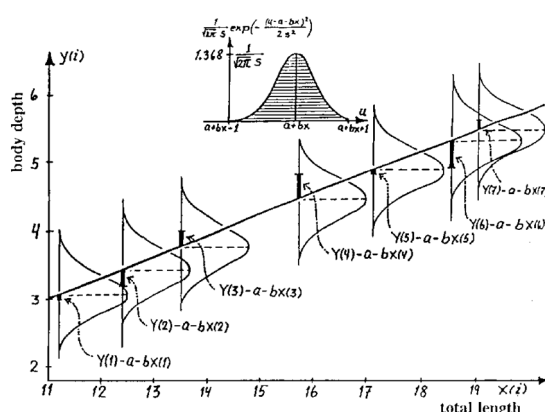
20

修正ポアソン回帰分析

- ▶ まさにその通り！（なので、なぜこれで、リスク比の推定がうまくいくのかがわかりにくい！）
- ▶ 種明かし：回帰モデルとは、もともと、「説明変数の組 $\mathbf{x}$ が与えられたときのアウトカム Y の条件付き平均 $E[Y|\mathbf{x}]$ 」のモデルのことである
- ▶ 回帰モデルとは、この条件付き平均が、1次関数のモデルで表されるというモデルのこと
 - ▶ $\log\{E(Y)\} = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$
- ▶ 我々が回帰分析で推定しているのは、この「平均」のみであって、「分布全体」ではない！

21

線形回帰モデルでも...



<https://mobiledevmemo.com/when-why-and-how-you-should-use-linear-regression/>

- ▶ 線形回帰モデルでは、誤差項の分布が正規分布に従うという仮定を置いている
- ▶ しかし、実際は、最尤法（普段、我々が使っている、線形回帰の推定方法）で求められる回帰係数の推定値は、**誤差項の分布が正規分布でなくても（誤っていても）、バイアスのない推定値になることが数学的に証明されている**
- ▶ 「平均」を正しく推定したいとき、「分布全体」の正しい仮定は必要ない。「平均」についてだけ正しい仮定を置いていけばよい。

22

参考 : Gauss-Markovの定理

- ▶ 線形回帰モデルにおいて、最小二乗法（正規分布を仮定したもとでの最尤法）から得られる推定量は、最良線形不偏推定量（Best Linear Unbiased Estimator; BLUE）となる
- ▶ 回帰係数の推定をバイアスなく行いたいのであれば、誤差項・アウトカムの正規分布の仮定はまったく必要ない
- ▶ 平均（回帰関数）の関数形さえ正しく仮定できていれば、左右非対称な歪んだ確率分布でも、2項分布のような離散分布であっても、回帰係数はバイアスなく推定できる（平均の推定の不偏性は、分布形の誤特定によって崩れないという頑健性がある）

23

修正ポアソン回帰分析

- ▶ ポアソン回帰でも、線形回帰モデルのGauss-Markovの定理の同様、平均（回帰関数）のモデルが正しければ、それ以外の仮定は間違えていても、回帰係数の推定値はバイアスのないものが得られる
 - ▶ $\log\{E(Y)\} = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$
- ▶ ざっくりと、回帰関数の関数形が「1次関数のモデルとなっており、変数の組が正しく選択されて」いれば、
- ▶ 真の分布が「2項分布」であっても、「ポアソン分布」のモデルをあてはめて、回帰係数（対数リスク比）はバイアスなく推定できる！

24

修正最小二乗回帰

- ▶ リスク差の回帰モデルも同様、真の「平均の関数モデル」が以下のものであるときに
 - ▶ $\Pr(Y = 1) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$
- ▶ 線形回帰モデル（最小二乗回帰）をあてはめると
 - ▶ $E(Y) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$
- ▶ 「平均の関数モデル」さえ正しければ、確率分布の型は間違えていても（2項分布→正規分布）、回帰係数はバイアスのない推定値が得られる
- ▶ これがそのままリスク差の推定値と解釈することができる！

Cheung (2007) 25

信頼区間の計算：ロバスト分散

- ▶ 一般的な回帰分析では、モデル分散という分散の推定公式を使って、信頼区間を計算している
- ▶ 修正ポアソン回帰・修正最小二乗回帰では、確率分布の型を敢えて誤ったものとして分析を行っているので、普通の分散の公式で正しい信頼区間を計算することができない
- ▶ 代わりに、ロバスト分散という、別な分散の公式を使えばよい（ここを置き換えるので、「修正」とされている）
- ▶ 普通のポアソン回帰・線形回帰の関数（Rのglmなど）で信頼区間を計算すると、間違えたものが出力されるので、絶対に使ってはいけない！

26

ロッテルダムの乳がん臨床研究

- ▶ Lampel et al. (2015) は、オランダのロッテルダムにおいて、乳がんの外科手術を受けた患者2,982人の予後を評価したコホート研究の結果を報告している
- ▶ ここでは、5年以内に再発を起こしたかどうかを2値アウトカムとしたデータ解析を行う
- ▶ 1,181人（39.6%）に対して、5年以内の再発イベントが観察されており、オッズ比によるリスク比の近似がうまくいかないと考えられる条件となっている

27

rqlm: Modified Poisson and Least-Squares Regressions for Binary Outcome

Modified Poisson and least-squares regression analyses for binary outcomes of Zou (2004) <[doi:10.1093/aje/kwh090](https://doi.org/10.1093/aje/kwh090)> and Cheung (2007) <[doi:10.1093/aje/kwm223](https://doi.org/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. Also, 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 Goshio (2024) <Forthcoming>).

Version: 1.2-1
Depends: R (≥ 3.5.0)
Imports: stats, MASS, sandwich
Published: 2024-01-19
Author: Hisashi Noma [aut, cre]
Maintainer: Hisashi Noma <noma at ism.ac.jp>
License: [GPL-3](https://www.gnu.org/licenses/gpl-3.0.html)
NeedsCompilation: no
Materials: [NEWS](#)
CRAN checks: [rqlm results](#)

Documentation:

Reference manual: [rqlm.pdf](#)

簡単なコマンドで、修正ポアソン回帰、修正最小二乗回帰を実行することができる R ライブラリ “rqlm” を作成しました。信頼区間の計算も、ロバスト分散のものになっています。

<https://cran.r-project.org/web/packages/rqlm/index.html>

28

事例コマンド

```
3  
4 # (1) ライブラリのインストール  
5  
6 install.packages("rqlm")  
7  
8  
9 # (2) ライブラリの読み込み  
10  
11 library("rqlm")  
12  
13  
14 # (3) データセットの読み込み  
15  
16 rot <- read.csv("JJSS2024.csv")  
17  
18  
19 # (4) アウトカム変数の分布が「2項分布」であると仮定した回帰分析  
20  
21 glm(y ~ size + nodes + grade + age + pgr, data=rot, family=binomial(link="logit")) # ロジスティック回帰分析  
22 glm(y ~ size + nodes + grade + age + pgr, data=rot, family=binomial(link="log")) # リスク比の2項回帰分析  
23 glm(y ~ size + nodes + grade + age + pgr, data=rot, family=binomial(link="identity")) # リスク差の2項回帰分析  
24
```

29

ロジスティック回帰分析

```
> glm(y ~ size + nodes + grade + age + pgr, data=rot, family=binomial(link="logit"))# ロジスティック回帰分析  
  
Call: glm(formula = y ~ size + nodes + grade + age + pgr, family = binomial(link = "logit"),  
  data = rot)  
  
Coefficients:  
(Intercept)      size>50      size20-50      nodes      grade      age      pgr  
   -1.54796      0.74075      0.52428      0.14898      0.48810     -0.01601     -0.03386  
  
Degrees of Freedom: 2981 Total (i.e. Null); 2975 Residual  
Null Deviance: 4004  
Residual Deviance: 3576 AIC: 3590
```

ロジスティック回帰分析は、オッズ比を推定する方法であり、基本的に、推定値が得られないということはない（まず確実にオッズ比の推定値は得られる）

30

リスク比・リスク差の2項回帰分析

```
> glm(y ~ size + nodes + grade + age + pgr, data=rot, family=binomial(link="log")) # リスク比の2項回帰分析
Error: no valid set of coefficients has been found: please supply starting values
> glm(y ~ size + nodes + grade + age + pgr, data=rot, family=binomial(link="identity")) # リスク差の2項回帰分析
Error: no valid set of coefficients has been found: please supply starting values
```

2項回帰によるリスク比・リスク差を推定する方法（対数リンク・恒等リンクによる2項回帰）は、解析プログラムが収束せず、推定値が定義できないことが多い。

この乳がんの臨床研究のケースでも、リスク比・リスク差は求められない。プログラム上では、このように、エラーが出る。

31

表 1 `rqlm` 関数において指定することのできるオプション。

引数	
<code>formula</code>	回帰式を指定するオブジェクト。アウトカム変数と説明変数の組を “~” でつないだ式とする（例： <code>y~x1+x2+x3</code> ）。 <code>lm</code> や <code>glm</code> の <code>formula</code> と同様のオブジェクトであり、スプライン関数などをモデル化することも可能。
<code>data</code>	解析対象となるデータセットのオブジェクトを指定する引数。
<code>family</code>	アウトカム変数の分布形（より厳密には、分散関数の型）を指定する引数。 <code>gaussian</code> とすれば修正最小二乗回帰となり、 <code>poisson</code> とすれば修正ポアソン回帰となる。 <code>glm</code> における <code>family</code> と同様、リンク関数を指定することも可能。
<code>eform</code>	回帰係数の推定値・信頼区間を指数関数で変換して出力するかどうかを指定する。TRUE であれば指数関数での変換を行い、FALSE であれば変換を行わない。
<code>ci</code>	信頼区間を計算する際の信頼係数を指定する。デフォルトは 0.95。
<code>digits</code>	分析結果の出力の小数点以下の桁数を指定する。デフォルトは 4 桁。

32

修正ポアソン回帰分析

```
> rqlm(y ~ size + nodes + grade + age + pgr, data=rot, family=poisson, eform=TRUE)
              exp(coef)      SE      CL      CU P-value
(Intercept)    0.1791 0.1956 0.1221 0.2628   0.000
size>50         1.5645 0.0678 1.3699 1.7867   0.000
size20-50       1.4228 0.0526 1.2834 1.5774   0.000
nodes           1.0515 0.0036 1.0441 1.0588   0.000
grade           1.3510 0.0601 1.2010 1.5198   0.000
age             0.9926 0.0017 0.9894 0.9958   0.000
pgr             0.9775 0.0114 0.9560 0.9995   0.045
```

修正ポアソン回帰分析。2項回帰（対数リンク）では、推定値が得られなかったのに、推定値はしっかりと得られている。これらが、バイアスのないリスク比の推定値と信頼区間になっている。エラーが出ることはほぼあり得ない。

33

修正最小二乗回帰分析

```
> rqlm(y ~ size + nodes + grade + age + pgr, data=rot, family=gaussian)
              coef      SE      CL      CU P-value
(Intercept)  0.1633 0.0622  0.0413 0.2853  0.0087
size>50       0.1689 0.0322  0.1057 0.2321  0.0000
size20-50     0.1159 0.0184  0.0798 0.1520  0.0000
nodes         0.0298 0.0022  0.0254 0.0342  0.0000
grade         0.1001 0.0187  0.0634 0.1367  0.0000
age          -0.0032 0.0007 -0.0045 -0.0019  0.0000
pgr          -0.0067 0.0031 -0.0128 -0.0005  0.0343
```

修正最小二乗回帰分析。2項回帰（恒等リンク）では、推定値が得られなかったのに、やはり、推定値は得られている。これらが、バイアスのないリスク差の推定値と信頼区間になる。こちらも、エラーが出ることはほぼあり得ない。

34

表 2 ロッテルダムにおける乳がん臨床研究データの多変量解析の結果。

	ロジスティック回帰	修正ポアソン回帰	修正最小二乗回帰
	オッズ比 (95%信頼区間)	リスク比 (95%信頼区間)	リスク差 (95%信頼区間)
腫瘍径 (vs. ≤ 20mm)			
21-50mm	2.098 (1.556, 2.827)	1.564 (1.370, 1.787)	0.169 (0.106, 0.232)
> 50mm	1.689 (1.426, 2.827)	1.423 (1.283, 1.577)	0.116 (0.080, 0.152)
リンパ節転移の数	1.161 (1.129, 1.193)	1.052 (1.044, 1.059)	0.030 (0.025, 0.034)
進行度	1.629 (1.350, 1.967)	1.351 (1.201, 1.519)	0.100 (0.063, 0.137)
年齢	0.984 (0.978, 0.990)	0.993 (0.989, 0.996)	-0.003 (-0.005, -0.002)
プロゲステロン受容体	0.967 (0.934, 1.001)	0.978 (0.956, 0.999)	-0.007 (-0.013, 0.000)

オッズ比は、この事例でも、リスク比を過大に評価してしまっている（バイアス大）。ロジスティック回帰は、こういうケースではNG。修正ポアソン回帰を使いましょう。

35

落穂拾い

- ▶ 臨床予測モデルの多変数予測モデルなどでは、効果の指標の推定が最終的な目的ではないため（正確な予測モデルを作ることが目的）、数学的な特性の優れたロジスティック回帰モデルがスタンダードな方法として用いられる
- ▶ 傾向スコアなどを使った因果推論の枠組みでは、別なアプローチによって、交絡によるバイアスを調整したリスク比・リスク差を推定することができる方法が提案されている（付加的な仮定を用いることで、セミパラメトリック有効な推定量を構成することもできる）

Steyerberg (2019), Imbens and Rubin (2015) 36

おわりに

- ▶ ロジスティック回帰分析は、長らく 2 値アウトカムの回帰分析のスタンダードな方法として使われてきたが、オッズ比は効果の指標としての直接的な解釈ができないため、リスク比・リスク差が推定できる条件下で、使う意義はそれほどない
- ▶ 簡便な計算方法で、リスク比・リスク差の多変量解析を行うことができる修正ポアソン回帰・修正最小二乗回帰は、今後、臨床研究・疫学研究の実践において、より普及していくものと思われる

37

文献

- ▶ Cheung, Y. B. (2007). A modified least-squares regression approach to the estimation of risk difference. *Am J Epidemiol* **166**, 1337-1344.
- ▶ Greenland, S. (1987). Interpretation and choice of effect measures in epidemiologic analysis. *Am J Epidemiol* **125**, 761-768.
- ▶ Imbens, G. W., and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge: Cambridge University Press.
- ▶ Lampelj, M., Arko, D., Cas-Sikosek, N., *et al.* (2015). Urokinase plasminogen activator (uPA) and plasminogen activator inhibitor type-1 (PAI-1) in breast cancer - correlation with traditional prognostic factors. *Radiol Oncol* **49**, 357-364.
- ▶ McNutt, L. A., Wu, C., Xue, X., and Hafner, J. P. (2003). Estimating the relative risk in cohort studies and clinical trials of common outcomes. *Am J Epidemiol* **157**, 940-943.
- ▶ Nurminen, M. (1995). To use or not to use the odds ratio in epidemiologic analyses. *Eur J Epidemiol* **11**, 365-371.

38

- ▶ Schulz, K. F., Altman, D. G., Moher, D., and Group, C. (2010). CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* **340**, c332.
- ▶ Steyerberg, E. W. (2019). *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*, 2nd edition. New York: Springer.
- ▶ Wallenstein, S., and Bodian, C. (1987). Epidemiologic programs for computers and calculators. Inferences on odds ratios, relative risks, and risk differences based on standard regression programs. *Am J Epidemiol* **126**, 346-355.
- ▶ Zou, G. (2004). A modified Poisson regression approach to prospective studies with binary data. *Am J Epidemiol* **159**, 702-706.

```
1  # (1) ライブラリのインストール
2
3  install.packages("rqml")
4
5
6  # (2) ライブラリの読み込み
7
8  library("rqml")
9
10
11 # (3) データセットの読み込み
12
13 rot <- read.csv("JJSS2024.csv")
14
15
16 # (4) アウトカム変数の分布が「2項分布」と仮定した回帰分析
17
18 glm(y ~ size + nodes + grade + age + pgr, data=rot, family=binomial(link="logit"))
19   # ロジスティック回帰分析
20 glm(y ~ size + nodes + grade + age + pgr, data=rot, family=binomial(link="log"))
21   # リスク比の2項回帰分析
22 glm(y ~ size + nodes + grade + age + pgr, data=rot, family=binomial(link="identity"))
23   # リスク差の2項回帰分析
24
25
26 # (5) 修正ポアソン回帰(リスク比のための回帰分析)
27
28 rqml(y ~ size + nodes + grade + age + pgr, data=rot, family=poisson, eform=TRUE)
29
30
31 # (6) 修正最小二乗回帰(リスク差のための回帰分析)
32
33 rqml(y ~ size + nodes + grade + age + pgr, data=rot, family=gaussian)
```