

# 混合効果モデルとマルチレベルモデル： 最新の国際医学誌のエビデンスを読み解くための 基礎知識と実践

野間 久史  
統計数理研究所

2024年7月14日

2024年度 LUNAレジストリー臨床研究合宿

e-mail: [noma@ism.ac.jp](mailto:noma@ism.ac.jp)

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

1

## 本日のお話

- ▶ マルチレベルモデル (multilevel model)
  - ▶ 相関を伴うデータの「クラスター構造」による統計的な不確実性を説明するための統計モデル
  - ▶ 経時的繰り返し測定データの解析, 多施設臨床研究, クラスターランダム化試験などの解析に使われる
- ▶ 線形混合効果モデル, ロジスティック混合効果モデル, ポアソン混合効果モデル, Frailtyモデル (生存時間解析)
  - ▶ これらすべてがマルチレベルモデル
  - ▶ 複数レベルのサンプリングモデルから構成される統計モデルの総称

2

## てんかんの臨床試験の事例

- ▶ Thall and Vail (1990) では、59名のてんかんの患者を対象として、progabideという薬剤の有効性を評価するためのプラセボ対照のランダム化比較試験を行っている
- ▶ ProgabideとPlaceboを、59名の患者にランダムに割り付け、8週間の間、各患者の発作の回数を記録した
- ▶ 2週間区分ごと（0-2, 2-4, 4-6, 6-8週間）の発作回数のデータが記録されており、それらを2群間で比較して、Progabideを使用することによって、各2週間のインターバル上での発作の平均回数が減るかどうかを評価したい

3

## データセットの構造

	A	B	C	D	E	F	G	H	I
	subject	trt	period	y	base	age	V4	lbase	lage
2	1	placebo	1	5	11	31	0	-0.75635	0.114204
3	1	placebo	2	3	11	31	0	-0.75635	0.114204
4	1	placebo	3	3	11	31	0	-0.75635	0.114204
5	1	placebo	4	3	11	31	1	-0.75635	0.114204
6	2	placebo	1	3	11	30	0	-0.75635	0.081414
7	2	placebo	2	5	11	30	0	-0.75635	0.081414
8	2	placebo	3	3	11	30	0	-0.75635	0.081414
9	2	placebo	4	3	11	30	1	-0.75635	0.081414
10	3	placebo	1	2	6	25	0	-1.36249	-0.10091
11	3	placebo	2	4	6	25	0	-1.36249	-0.10091
12	3	placebo	3	0	6	25	0	-1.36249	-0.10091
13	3	placebo	4	5	6	25	1	-1.36249	-0.10091
14	4	placebo	1	4	8	36	0	-1.07481	0.263735
15	4	placebo	2	4	8	36	0	-1.07481	0.263735
16	4	placebo	3	1	8	36	0	-1.07481	0.263735
17	4	placebo	4	4	8	36	1	-1.07481	0.263735
18	5	placebo	1	7	66	22	0	1.035406	-0.22874
19	5	placebo	2	18	66	22	0	1.035406	-0.22874
20	5	placebo	3	9	66	22	0	1.035406	-0.22874
21	5	placebo	4	21	66	22	1	1.035406	-0.22874
22	6	placebo	1	5	27	29	0	0.141588	0.047512
23	6	placebo	2	2	27	29	0	0.141588	0.047512
24	6	placebo	3	8	27	29	0	0.141588	0.047512
25	6	placebo	4	7	27	29	1	0.141588	0.047512

個人ごとに、4つの時間区分に分けて、2週間ごとの発作回数のデータが記録されている

一般的なポアソン回帰などで解析をするケースでは、「1人あたり1レコード」の形式で、データが記録されているが、「1人あたり4レコード」のデータが記録されている。

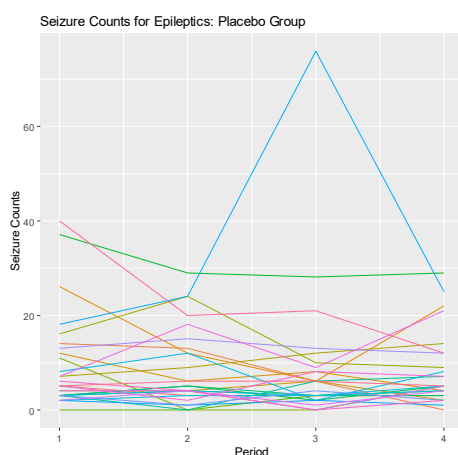
4

## ポアソン回帰による統計解析

- ▶ 「発作の回数」というカウントデータがアウトカムなので、ポアソン回帰で解析をすればいいですよね？
- ▶ ちょっと待った！！
- ▶ ポアソン回帰をはじめとする、我々が臨床研究の統計解析で使用する基本モデル（線形回帰，ロジスティック回帰，ポアソン回帰，Cox回帰など）は、すべて「対象となるデータは**独立**」という仮定を用いている
- ▶ 独立性（independence）：お互いの取る値はまったく無関係で、影響を及ぼし合うことはない

5

## 患者ごとの発作回数のプロフィール



- ▶ 個人ごとに、発作回数のプロフィールは異なる
- ▶ 「もともと発作回数の多い患者」と「もともと発作回数の少ない患者」が混在している
- ▶ 異なる患者の発作回数のデータは、互いに影響を及ぼし合うことはないので、独立と考えられるが...
- ▶ 同一個人の異なる時点の発作回数のデータは、独立ではないのでは？

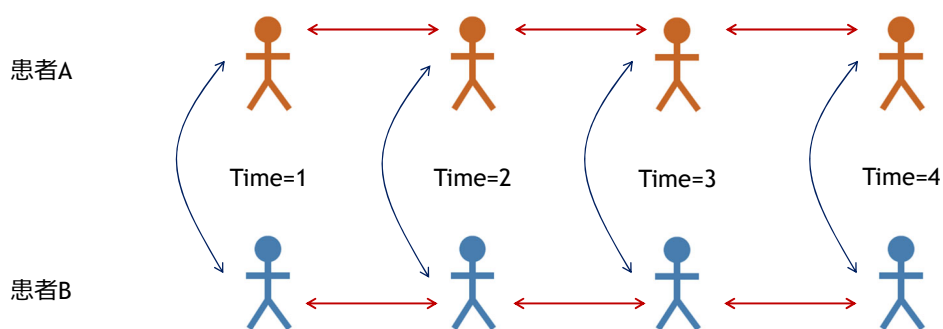
6

## 経時的繰り返し測定データ

- ▶ 同一個人に対して、繰り返し測定されるアウトカムのデータは、一般的に「独立」と見なすことができず、「相関」を持つことが一般的
- ▶ 我々が普段使っている、基本的な統計解析の方法は、すべての対象となるデータが「独立」とであると仮定している
  - ▶ カイ二乗検定, t検定, ロジランク検定
  - ▶ 線形回帰, ロジスティック回帰, ポアソン回帰, Cox回帰
- ▶ 「相関」を無視した解析を行ってしまうと、通常、信頼区間やP値が誤ったものになってしまうことが知られている！

7

## 繰り返し測定データにおける「相関」



①異なる個人のデータは、互いに独立（まったく無関係）

②同一個人の繰り返し測定されるデータは、互いに影響を及ぼし合う（独立でない）

8

## Notation

- ▶  $Y_{it}$ :  $i$  番目の患者の時点  $t$  のアウトカム
- ▶  $x_{it}$ :  $i$  番目の患者の時点  $t$  の治療を表す変数
  - ▶  $= 1$  (progabide),  $= 0$  (placebo)
- ▶  $z_{1,it}, z_{2,it}$ :  $i$  番目の患者の時点  $t$  の調整変数
  - ▶ 簡単のため、調整変数が2つのケースとしますが、いくつかの場合でも、以降の議論は同じです

9

## ポアソン回帰モデル

- ▶ カウントデータや率 (incidence rate) のデータの解析に使われる回帰モデル
- ▶  $\log\{E[Y_{it}]\} = \beta_0 + \beta_1 x_{it} + \beta_2 z_{1,it} + \beta_3 z_{2,it}$ 
  - ▶ 各患者・各時点のイベントの回数の平均の対数値が、1次関数で表されるという回帰関数のモデル
  - ▶ すべてのアウトカムが「独立」であれば、回帰係数の正しい推定値・信頼区間が得られる
- ▶ しかし、独立性が成り立たないケースでは、誤った結果が得られてしまう！

10

## 「相関」をどのようにモデル化する？

- ▶ てんかんの臨床試験の事例では、個人ごとに発作の回数の全般的な水準（ベースラインの水準）が異なっていた
- ▶  $\log\{E[Y_{it}]\} = \beta_0 + \gamma_i + \beta_1 x_{it} + \beta_2 z_{1,it} + \beta_3 z_{2,it}$

このように、個人ごとの全体としての平均の大小の水準（ベースラインレベル）を表す項を加えるのが妥当では？

- ▶ 実は、相関のあるクラスタごとに、回帰関数の値の大小が異なる「クラスタ効果」を表す項を加えることで、回帰モデルの中での相関を説明することができる！

11

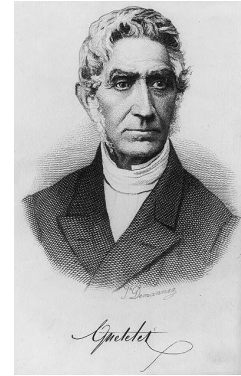
## クラスタ効果 $\gamma_i$ のモデル化

- ▶  $\log\{E[Y_{it}]\} = \beta_0 + \gamma_i + \beta_1 x_{it} + \beta_2 z_{1,it} + \beta_3 z_{2,it}$ 
  - ▶ クラスタ効果を表すパラメータ  $\gamma_i$  は、全員で値が異なっていると仮定して、データから推定しようとする、人数分の回帰係数（59個）が必要となる
  - ▶ 関心のあるパラメータ  $\beta_1$  の推定精度が著しく低下する or そもそも推定することができなくなってしまう
- ▶  $\gamma_i$  を確率変数（確率的に変動する数値）と見なして、正規分布に従うと仮定してモデル化する方法が考案された
  - ▶  $\gamma_i \sim N(0, \sigma^2)$
  - ▶ 推定すべきパラメータは  $\sigma^2$  1つに集約される！

12

## Adolphe Quetelet (1796-1874)

- ▶ ベルギーの社会学者、数学者、天文学者、統計学者
- ▶ BMI (Body Mass Index) を提案した人として有名
- ▶ 犯罪率、結婚率、自殺率などの人間社会におけるさまざまな指標の統計学的な法則を明らかにすることを目標として、社会科学における統計的方法の導入を先導した（近代統計学の父とも）
- ▶ ケトレーは、社会の中で、人の持つさまざまな特徴は、**正規分布のような分布に従う法則がある**として、その中心に位置し平均的特徴を持つ人を「平均人（l'homme moyen）」と呼んだ



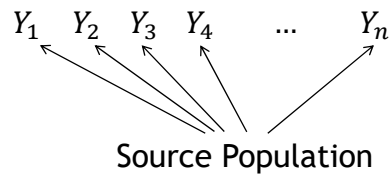
[https://en.wikipedia.org/wiki/Adolphe\\_Quetelet](https://en.wikipedia.org/wiki/Adolphe_Quetelet) 13

## ポアソン混合効果モデル

- ▶  $\log\{E[Y_{it}]\} = \beta_0 + \gamma_i + \beta_1 x_{it} + \beta_2 z_{1,it} + \beta_3 z_{2,it}$ 
  - ▶  $\gamma_i \sim N(0, \sigma^2)$  **変量効果 (random effects) という**
- ▶  $\alpha_i = \beta_0 + \gamma_i$  とすると、切片が変量効果となった回帰モデルと見なすこともできる ( $\alpha_i \sim N(\beta_0, \sigma^2)$ )
  - ▶ **ランダム切片モデル (random intercept model)**
- ▶ 変量効果でない回帰係数を固定効果 (fixed effects) という
- ▶ **固定効果・変量効果が混合したモデルとして、混合効果モデル (mixed effects model) と呼ばれる**
- ▶ 相関を適切に説明した  $\beta_1, \beta_2, \beta_3$  の推定値・信頼区間を得ることができる

14

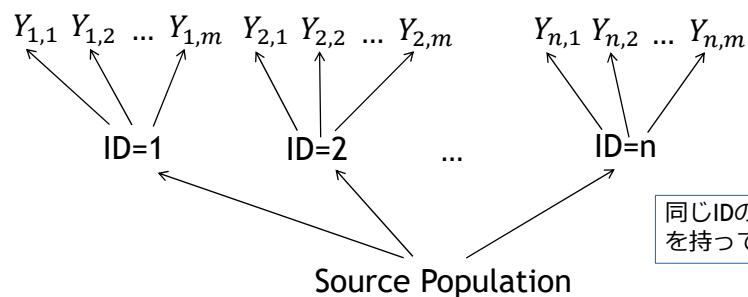
## ポアソン回帰モデルで想定している仮定



1つ1つのデータは、独立に、まったく同じ分布に従って得られるものであると仮定している

15

## ポアソン混合効果モデルでは



クラスター構造を持つデータを、階層型の構造を持つモデルでモデル化したもの (= マルチレベルモデル)

16



## ここまでのまとめ

- ▶ 経時的繰り返し測定データなど、特定のデータの「クラスタ」に相関があるデータセットでは、相関を適切に説明する方法を使わなくては、誤った解析結果が得られる
- ▶ 相関は、回帰モデルの中で、クラスタごとに、共通のベースライン効果（変量効果）を加えてやることでモデル化することができる（＝混合効果モデル，マルチレベルモデル）
- ▶ 混合効果モデルを用いることによって、すべての回帰係数に、相関の影響を適切に考慮した推定値・信頼区間を得ることができる

17

## R言語 R language

- ▶ オープンソース・フリーソフトウェアの統計解析のためのプログラミング言語およびその開発環境
- ▶ AT&Tベル研究所が開発した S言語をもとにしている
- ▶ 臨床研究で用いられるスタンダードな解析手法はほぼすべて、S言語を母体とするモジュールとなっており、信頼のできるソフトウェアとして、広く医学論文でも用いられている
- ▶ GPL3ライセンスに準拠しており、すべての機能を無償で利用することができる（付属するパッケージも同様）



CRAN (The Comprehensive R Archive Network) <https://cran.r-project.org/>

18

## ライブラリのインストール

- ▶ 本日の実習で使用するライブラリのインストール方法です

```
install.packages("lme4")           # 混合効果モデルによる解析を行うためのパッケージです  
install.packages("openxlsx")       # Excelのファイルを読み込むためのパッケージです
```

- ▶ Rで、最も多く使われているマルチレベルモデルを用いた解析のためのパッケージは、lme4 です
- ▶ このうち、lmer, glmer 関数を使います

19

## データセットの読み込み

- ▶ 事前に配布したExcelファイルから、てんかんの臨床試験のデータを読み込みましょう

```
library("lme4")  
library("openxlsx")  
  
setwd("D:¥¥Dropbox¥¥Share")  
  
epil <- read.xlsx("example20240714.xlsx", sheet="epil") # データセットの読み込み  
head(epil, 30)
```

20

## ポアソン回帰分析

```
> gml <- glm(y ~ trt + lbase + lage, family=poisson, data=epil)
> summary(gml)

Call:
glm(formula = y ~ trt + lbase + lage, family = poisson, data = epil)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.1910  -1.5011  -0.2793   0.7582  10.6552

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.70874    0.04080  41.877 < 2e-16 ***
trtprogabide -0.01685    0.04820  -0.350   0.727
lbase        1.22422    0.03253  37.632 < 2e-16 ***
lage         0.57882    0.10998   5.263 1.42e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 2517.83  on 235  degrees of freedom
Residual deviance:  954.77  on 232  degrees of freedom
AIC: 1728.7

Number of Fisher Scoring iterations: 5
```

同一個人に繰り返し測定されるアウトカムの相関を無視したポアソン回帰分析の結果です。特に、標準誤差 (Std. Error) の推定値が誤ったものとなることが多いです。

21

## glmerによるポアソン混合効果モデル

```
► glmer(y ~ trt + lbase + lage + (1|subject),
family=poisson, nAGQ=25, data=epil)
```

回帰式の指定は、lmやglmと同じ。

glmerという関数を使う

(1|subject) で、変量効果を「subject」という変数の単位で入れることを指定できます。クラスタ効果の「単位」を表す変数を、ここで指定します。

近似計算の精度を指定する変数です。nAGQ=25としておきましょう。

family=poissonとすることで、ポアソン混合効果モデルでの解析を行うことができます

22

## ポアソン混合効果モデルによる解析

```
> gm2 <- glmer(y ~ trt + lbase + lage + (1|subject), family=poisson, nAGQ = 25, data=epil)
> summary(gm2)
Generalized linear mixed model fit by maximum likelihood (Adaptive Gauss-Hermite Quadrature, nAGQ = 25) ['glmerMod']
Family: poisson ( log )
Formula: y ~ trt + lbase + lage + (1 | subject)
Data: epil

           AIC      BIC    logLik deviance df.resid 
      586.4      603.8    -288.2     576.4      231 

Scaled residuals:
      Min       1Q   Median       3Q      Max 
-3.1361 -0.9187 -0.1165  0.5880  6.9437 

Random effects:
 Groups Name      Variance Std.Dev.
 subject (Intercept) 0.2677   0.5174
Number of obs: 236, groups: subject, 59

Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    1.7937     0.1075   16.685  <2e-16 ***
trtprogabide   -0.3153     0.1511   -2.087   0.0369 *
lbase          1.0273     0.1015   10.120  <2e-16 ***
lage           0.3318     0.3440    0.965   0.3348
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
      (Intr) trtprg lbase
trtprogabide -0.712
lbase        -0.086 -0.039
lage         -0.095  0.125  0.173
```

同一個人に繰り返し測定されるアウトカムの相関を適切に調整したポアソン混合効果モデルによる解析の結果です。

23

## 相対リスクと95%信頼区間の計算

```
> coeff(gm2, eform=TRUE)
              exp(coef)      SE      CL      CU P-value
(Intercept)    6.0119 0.1075 4.8697 7.4220 0.0000
trtprogabide    0.7295 0.1511 0.5425 0.9810 0.0369
lbase           2.7934 0.1015 2.2894 3.4083 0.0000
lage            1.3935 0.3440 0.7100 2.7348 0.3348
```

ポアソン回帰分析と同様、ポアソン混合効果モデルによる解析の結果も、対数変換されたスケールでの相対リスクの推定値が算出されるので、指数関数で変換して報告することが一般的です。指数関数での変換と信頼区間の計算は、こちらのコマンドでまとめて実行できます。

24

## よく使われる混合効果モデル

- ▶ 同様に、アウトカム変数の型ごとに使われる基本モデルは異なるが、クラスタごとの変量効果を回帰モデルに組み込むだけで、相関を調整した解析を行うことができる
- ▶ 線形回帰モデル→線形混合効果モデル（連続）
- ▶ ロジスティック回帰モデル→ロジスティック混合効果モデル（2値）
- ▶ ポアソン回帰モデル→ポアソン混合効果モデル（率・カウント）
- ▶ Cox回帰モデル→Frailtyモデル（生存時間）

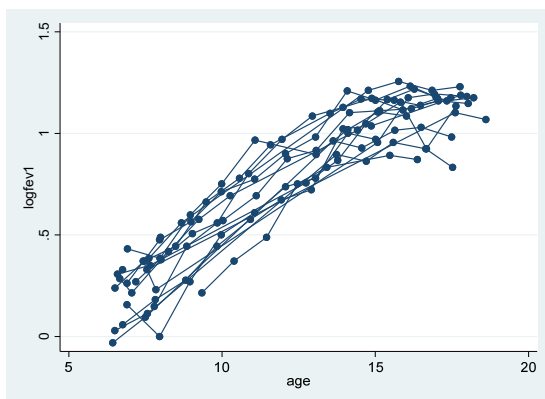
25

## Dataset: fev1

- ▶ The Six Cities Study of Air Pollution and Health was a longitudinal study designed to characterize lung growth as measured by changes in pulmonary function in children and adolescents, and the factors that influence lung function growth. A cohort of 13,379 children born on or after 1967 was enrolled in six communities across the U.S.: Watertown (Massachusetts), Kingston and Harriman (Tennessee), a section of St. Louis (Missouri), Steubenville (Ohio), Portage (Wisconsin), and Topeka (Kansas). Most children were enrolled in the first or second grade (between the ages of six and seven) and measurements of study participants were obtained annually until graduation from high school or loss to follow-up. At each annual examination, spirometry, the measurement of pulmonary function, was performed and a respiratory health questionnaire was completed by a parent or guardian.
- ▶ The dataset contains a subset of the pulmonary function data collected in the Six Cities Study. The data consist of all measurements of FEV1, height and age obtained from a randomly selected subset of the female participants living in Topeka, Kansas. The random sample consists of 300 girls, with a minimum of one and a maximum of twelve observations over time.

<https://content.sph.harvard.edu/fitzmaur/ala2e/fev1.txt> 26

## FEV1と年齢の経時的測定データ



年齢と $\log(\text{FEV1})$ は、おおよそ同程度の傾きを持つ直線傾向の関数関係にありそうだが、個人ごとにベースラインのレベル（切片）は異なるようである

連続変数でも、同一個人の複数時点での繰り返し測定データは相関を持つため、適切な調整を行わなくてはならない（普通の線形回帰ではダメ！）。

27

## データセットの読み込み

- ▶ 事前に配布したExcelファイルから、6都市研究のデータを読み込みましょう

```
fev1 <- read.xlsx("example20240714.xlsx", sheet="fev1") # データセットの読み込み
head(fev1, 30)
```

28

## 線形回帰分析

```
> lml <- lm(logfevl ~ age + baseage + baseht, data=fevl)
> summary(lml)

Call:
lm(formula = logfevl ~ age + baseage + baseht, data = fevl)

Residuals:
    Min       1Q   Median       3Q      Max
-1.03438 -0.08665  0.00328  0.08929  0.44881

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.6359135   0.0540630   -30.26  <2e-16 ***
age           0.0860437   0.0009468    90.88  <2e-16 ***
baseage      -0.0818554   0.0041540   -19.70  <2e-16 ***
baseht       1.5892423   0.0589590    26.95  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1387 on 1990 degrees of freedom
Multiple R-squared:  0.8241,    Adjusted R-squared:  0.8239
F-statistic: 3109 on 3 and 1990 DF,  p-value: < 2.2e-16
```

同一個人に繰り返し測定されるアウトカムの相関を無視した線形回帰分析の結果です。特に、標準誤差 (Std. Error) の推定値が誤ったものとなることが多いです。

29

## lmerによる線形混合効果モデル

```
► lmer(logfevl ~ age + baseage + baseht + (1|id),
data=fevl)
```

lmerという関数を使う

回帰式の指定は、lmやglmと同じ。

(1|id) で、変量効果を「id」という変数の単位で入れることを指定できます。  
クラスタ効果の「単位」を表す変数を、ここで指定します。

lmerは、線形混合効果モデルの解析のみに使う関数なので、familyという引数はありません。

30

## 線形混合効果モデルによる解析

```
> lm2 <- lmer(logfevl ~ age + baseage + baseht + (1|id), data=fevl)
> summary(lm2)
Linear mixed model fit by REML ['lmerMod']
Formula: logfevl ~ age + baseage + baseht + (1 | id)
Data: fevl

REML criterion at convergence: -2875.5

Scaled residuals:
    Min       1Q   Median       3Q      Max
-5.2220 -0.5649  0.0353  0.6516  2.9150

Random effects:
Groups Name Variance Std.Dev.
id      (Intercept) 0.009329 0.09658
Residual              0.010470 0.10232
Number of obs: 1994, groups: id, 300

Fixed effects:
              Estimate Std. Error t value
(Intercept) -1.727488    0.103899  -16.63
age           0.085769    0.000740   115.91
baseage      -0.091320    0.008143   -11.21
baseht        1.717584    0.117283    14.64

Correlation of Fixed Effects:
      (Intr) age  baseag
age      -0.052
baseage  0.542 -0.029
baseht  -0.924 -0.010 -0.816
```

同一個人に繰り返し測定されるアウトカムの相関を適切に調整した線形混合効果モデルによる解析の結果です。

31

## 推定値と95%信頼区間の計算

```
> coeff(lm2)
              coef      SE      CL      CU P-value
(Intercept) -1.7275 0.1039 -1.9311 -1.5238      0
age           0.0858 0.0007  0.0843  0.0872      0
baseage      -0.0913 0.0081 -0.1073 -0.0754      0
baseht        1.7176 0.1173  1.4877  1.9475      0
```

線形混合効果モデルによる解析では、回帰係数が、そのまま、線形回帰モデルと同様、「傾き」もしくは「平均値の差」として解釈することができます。信頼区間の計算は、こちらのコマンドで実行できます。

32



ohio {geepack}

R Documentation

Ohio Children Wheeze Status

#### Description

The `ohio` data frame has 2148 rows and 4 columns. The dataset is a subset of the six-city study, a longitudinal study of the health effects of air pollution.

#### Usage

`ohio`

#### Format

This data frame contains the following columns:

`resp`

an indicator of wheeze status (1=yes, 0=no)

`id`

a numeric vector for subject id

`age`

a numeric vector of age, 0 is 9 years old

`smoke`

an indicator of maternal smoking at the first year of the study

#### References

Fitzmaurice, G.M. and Laird, N.M. (1993) A likelihood-based method for analyzing longitudinal binary responses, *Biometrika* **80**: 141–151.

<https://rdrr.io/cran/geepack/man/ohio.html>

33

## データセットの読み込み

- ▶ 事前に配布したExcelファイルから、同じく6都市研究のデータを読み込みましょう

```
ohio <- read.xlsx("example20240714.xlsx", sheet=" ohio ") # データセットの読み込み
head(ohio,30)
```

34

## ロジスティック回帰分析

```
> gm3 <- glm(resp ~ age + smoke, family=binomial, data=ohio)
> summary(gm3)

Call:
glm(formula = resp ~ age + smoke, family = binomial, data = ohio)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.6685  -0.5909  -0.5608  -0.5045   2.0613

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.88373    0.08384  -22.467  <2e-16 ***
age          -0.11341    0.05408   -2.097   0.0360 *
smoke         0.27214    0.12347    2.204   0.0275 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1829.1  on 2147  degrees of freedom
Residual deviance: 1819.9  on 2145  degrees of freedom
AIC: 1825.9

Number of Fisher Scoring iterations: 4
```

同一個人に繰り返し測定されるアウトカムの相関を無視したロジスティック回帰分析の結果です。特に、標準誤差 (Std. Error) の推定値が誤ったものとなることが多いです。

35

## glmerによるロジスティック混合効果モデル

```
► glmer(resp ~ age + smoke + (1|id),
  family=binomial , nAGQ=25, data=ohio)
```

回帰式の指定は、lmやglmと同じ。

(1|id) で、変量効果を「id」という変数の単位で入れることを指定できます。クラスタ効果の「単位」を表す変数を、ここで指定します。

glmerという関数を使う

近似計算の精度を指定する変数です。nAGQ=25としておきましょう。

family=binomialとすることで、ロジスティック混合効果モデルでの解析を行うことができます

36

## ロジスティック混合効果モデルによる解析

```
> gm4 <- glmer(resp ~ age + smoke + (1|id), nAGQ=25, family=binomial, data=ohio)
> summary(gm4)
Generalized linear mixed model fit by maximum likelihood (Adaptive Gauss-Hermite Quadrature, nAGQ = 25) ['glmerMod']
Family: binomial ( logit )
Formula: resp ~ age + smoke + (1 | id)
Data: ohio

      AIC      BIC    logLik deviance df.resid
 1603.3  1626.0   -797.6   1595.3     2144

Scaled residuals:
   Min       1Q   Median       3Q      Max
-1.3730 -0.2011 -0.1773 -0.1487  2.5076

Random effects:
 Groups Name      Variance Std.Dev.
 id      (Intercept) 4.687    2.165
Number of obs: 2148, groups: id, 537

Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.10153    0.21904  -14.159  < 2e-16 ***
age           -0.17563    0.06768   -2.595  0.00946 **
smoke          0.39857    0.27308    1.460  0.14442
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
      (Intr) age
age      0.244
smoke   -0.493 -0.008
```

同一個人に繰り返し測定されるアウトカムの相関を適切に調整したロジスティック混合効果モデルによる解析の結果です。

37

## オッズ比と95%信頼区間の計算

```
> coeff(gm4, eform=TRUE)
              exp(coef)      SE      CL      CU P-value
(Intercept)  0.0450 0.2190 0.0293 0.0691 0.0000
age           0.8389 0.0677 0.7347 0.9579 0.0095
smoke         1.4897 0.2731 0.8723 2.5442 0.1444
```

ロジスティック回帰分析と同様、ロジスティック混合効果モデルによる解析の結果も、対数変換されたスケールでのオッズ比の推定値が算出されるので、指数関数で変換して報告することが一般的です。指数関数での変換と信頼区間の計算は、こちらのコマンドでまとめて実行できます。

38

## 落穂拾い①

- ▶ 今回の解説では、切片だけが変量効果となる「ランダム切片モデル」のみについて解説を行いました。傾きも同様に変量効果とすることができます。多施設共同研究などで、治療効果が施設によって異なることが想定される場合などに、その異質性をモデル化することができます。
- ▶ Cox回帰に変量効果を加えたモデルは、歴史的経緯から、Frailtyモデルという呼び名が一般的です。Rでは、coxph関数の回帰式に、frailtyという関数にかけた項を加えることで解析ができます。

39

## 落穂拾い②

- ▶ クラスター効果による相関を調整するためのもう一つの方法として、一般化推定方程式（generalized estimating equation GEE）があります。GEEでは、アウトカム間の相関を、作業相関係数行列として、直接モデル化します。
- ▶ 経時的繰り返し測定データなどで、MAR (missing at random) で起こるアウトカムの欠測によるバイアスは、混合効果モデルを使った解析では自動的に調整されます（Rubinの無視可能性）。GEEでは、IPW法などを使わないと調整することができません（が、GEEのほうが緩い仮定のもとでのモデルを使っていますので、一長一短あります）。

40

## おわりに

- ▶ マルチレベルモデルを用いることで、アウトカム間の相関を適切に調整した統計解析を行うことができる
- ▶ 経時的繰り返し測定データ, 多施設共同研究, クラスターランダム化比較試験など、クラスター効果を考慮した解析が必要な研究で有用な方法となる
- ▶ Rなどのフリーの統計解析ソフトウェアでも、簡単なコマンドで実行することができるようになっているので、必要に応じて、有効活用していきましょう

41

## 文献

- ▶ Thall, P. F. and Vail, S. C. (1990) Some covariance models for longitudinal count data with over-dispersion. *Biometrics* **46**, 657-671.
- ▶ Fitzmaurice, G.M. and Laird, N.M. (1993) A likelihood-based method for analyzing longitudinal binary responses, *Biometrika* **80**: 141-151.

42

```
1  #-----
2  #
3  #   2024年度 LUNAレジストリー臨床研究合宿
4  #   2024年7月14日
5  #
6  #   「臨床研究・疫学研究におけるマルチレベルモデルを用いた統計解析」
7  #   統計数理研究所 野間久史
8  #
9  #   ※ 本プログラム・データセットの流用・転用など、二次利用はご遠慮ください。
10 #
11 #-----
12
13
14
15
16 #-----
17 # 1. パッケージのインストール
18 #-----
19
20 install.packages("lme4")           # 混合効果モデルによる解析を行うためのパッケージです
21 install.packages("openxlsx")       # Excelのファイルを読み込むためのパッケージです
22 install.packages("rqlm")           # 信頼区間・P値を一括計算するための関数が入っているパッケージです
23
24 # 1度インストールをしたら、このコマンドは飛ばしていただいて構いません。
25
26
27
28
29
30
31 #-----
32 # 2. ポアソン混合効果モデル
33 #-----
34
35 library("lme4")
36 library("openxlsx")
37 library("rqlm")
38
39 setwd("D:\\Dropbox\\Share")
40
41 epil <- read.xlsx("example20240714.xlsx", sheet="epil")      # データセットの読み込み
42 head(epil, 30)
43
44 gm1 <- glm(y ~ trt + lbase + lage, family=poisson, data=epil)
45 summary(gm1)
46 # 普通のポアソン回帰分析です。同一個人内で、繰り返し測定されたアウトカムの相関を無視した解析をしています。
47
48 gm2 <- glmer(y ~ trt + lbase + lage + (1|subject), family=poisson, nAGQ=25, data=epil)
49 summary(gm2)
50 #
51 # ポアソン混合効果モデルによる解析です。マルチレベルモデルによって、同一個人内で、繰り返し測定されたアウトカムの相関を調整した解析をしています。
52
53 coeff(gm2, eform=TRUE)
54 # ポアソン混合効果モデルによる回帰係数から推定した、相対リスクの推定値と95%信頼区間です。
55
56
57
58
59
60 #-----
61 # 3. 線形混合効果モデル
62 #-----
63
64 fev1 <- read.xlsx("example20240714.xlsx", sheet="fev1")      # データセットの読み込み
65 head(fev1, 30)
66
```

```
67  lm1 <- lm(logfev1 ~ age + baseage + baseht, data=fev1)
68  summary(lm1)
69  # 普通の線形回帰分析です。同一個人内で、繰り返し測定されたアウトカムの相関を無視した解析をしています。
70
71  lm2 <- lmer(logfev1 ~ age + baseage + baseht + (1|id), data=fev1)
72  summary(lm2)
73  #
74  # 線形混合効果モデルによる解析です。マルチレベルモデルによって、同一個人内で、繰り返し測定されたアウトカムの相関を調整した解析をしています。
75
76  coeff(lm2)
77  # 線形混合効果モデルによる回帰係数から推定した、回帰係数の推定値と95%信頼区間です。
78
79
80
81
82
83  #-----
84  # 4. ロジスティック混合効果モデル
85  #-----
86
87  ohio <- read.xlsx("example20240714.xlsx", sheet="ohio")      # データセットの読み込み
88  head(ohio, 30)
89
90  gm3 <- glm(resp ~ age + smoke, family=binomial, data=ohio)
91  summary(gm3)
92  # 普通のロジスティック回帰分析です。同一個人内で、繰り返し測定されたアウトカムの相関を無視した解析をしています。
93
94  gm4 <- glmer(resp ~ age + smoke + (1|id), nAGQ=25, family=binomial, data=ohio)
95  summary(gm4)
96  #
97  # ロジスティック混合効果モデルによる解析です。マルチレベルモデルによって、同一個人内で、繰り返し測定されたアウトカムの相関を調整した解析をしています。
98
99  coeff(gm4, eform=TRUE)
100 # ロジスティック混合効果モデルによる回帰係数から推定した、オッズ比の推定値と95%信頼区間です。
101
```