Multilevel Logistic Regression for Binary Data – Part 2

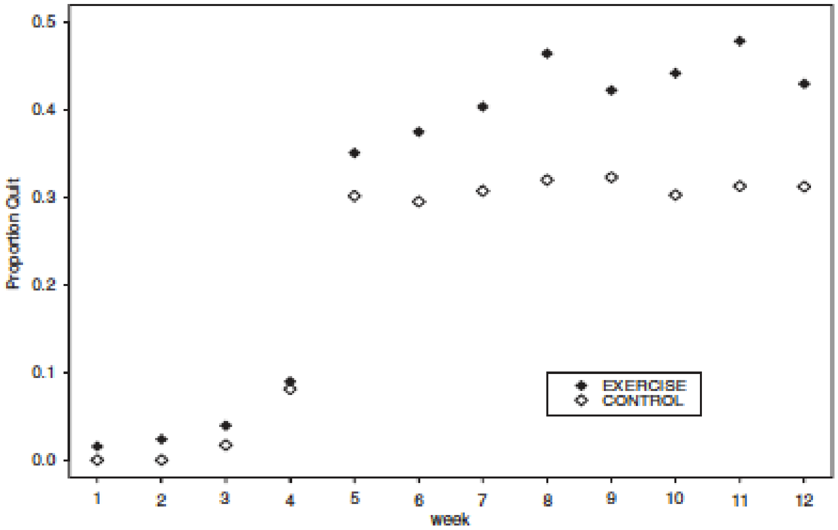
Nambari Short Course

16 July 2019

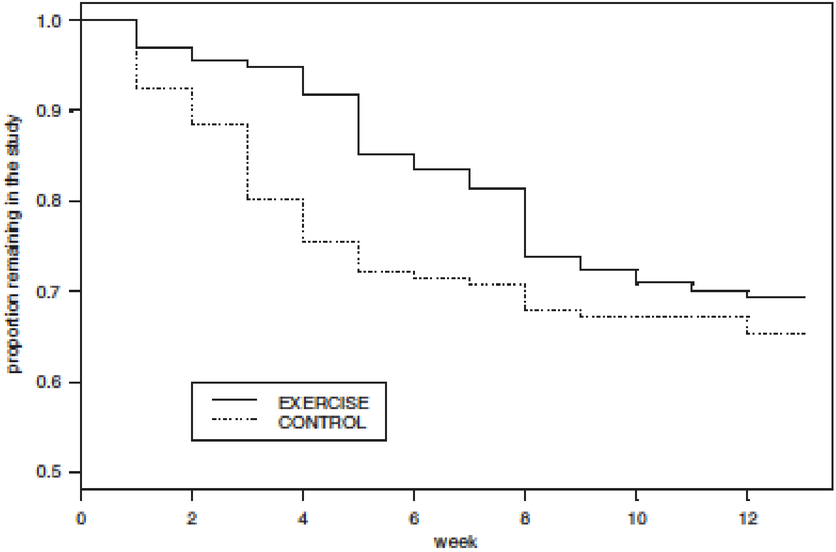
Overview

- Revisit data structure of CTQ data
- Model treatment effect using multilevel logit model
- Compare conditional (subject-specific) versus marginal (population-averaged) treatment effect

Smoking Cessation Study: Summaries



Smoking Cessation Study: Summaries



Data Analyses

1. Multilevel logistic regression with intercept only
2. Multilevel logistic regression on single covariate
3. MLR of time trend and treatment effect

Variables used in this analysis

$$
\begin{aligned}
Y_{ij} &= \text{quit status for person } i \text{ at time } t_j \\
&= 1 \text{ if quit, } 0 \text{ if not} \\
t_j &= \text{measurement time in weeks} \\
Z_i &= \text{treatment group } (1 = \text{exercise, } 0 = \text{control}) \\
X_i &= \text{baseline level of nicotine dependence } (0 \text{ to } 10)
\end{aligned}
$$

Data excerpt

	ID	week	wk>4	Z	X	Y
[1,]	305	4	0	0	8	0
[2,]	305	5	1	0	8	1
[3,]	305	6	1	0	8	1
[4,]	305	7	1	0	8	1
[5,]	305	8	1	0	8	1
[6,]	305	9	1	0	8	1
[7,]	305	10	1	0	8	1
[8,]	305	11	1	0	8	1
[9,]	305	12	1	0	8	1
[10,]	309	4	0	1	6	0
[11,]	309	5	1	1	6	0
[12,]	309	6	1	1	6	0
[13,]	309	7	1	1	6	0
[14,]	309	8	1	1	6	0
[15,]	309	9	1	1	6	0
[16,]	309	10	1	1	6	0
[17,]	309	11	1	1	6	0
[18,]	309	12	1	1	6	0

Model 2: Include individual-level covariate

Covariate here is nicotine dependence score (0 to 10)

Level 1

$$\begin{aligned} Y_{ij} &\sim \text{Ber}(\pi_{ij}) \\ \text{logit}(\pi_{ij}) &= \alpha_i + \beta X_i \end{aligned}$$

Level 2

$$\alpha_i \sim N(0, \tau^2)$$

Interpretation:

- Coefficient β is the *subject-specific* effect of X; i.e., the effect of X within an individual.
- Another interpretation is that it is the *conditional* effect of X (conditioning on α_i, the individual-level propensity to quit smoking)

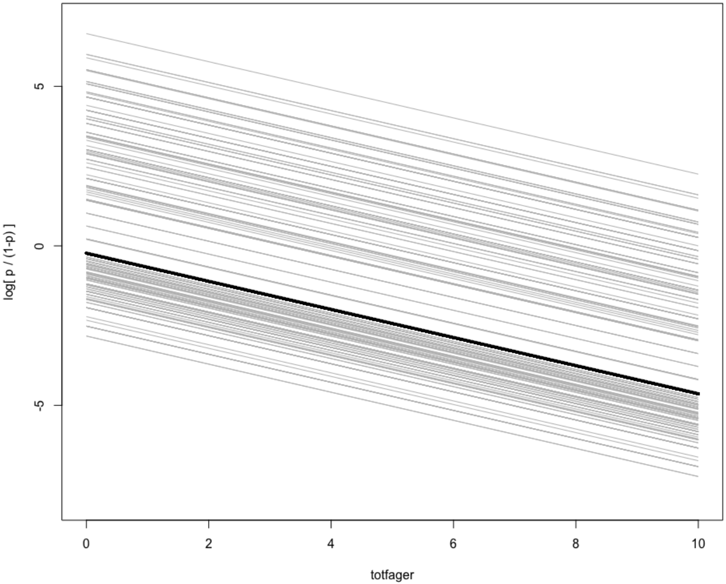
Fitting the model in R

```
> M1 = glmer( Y ~ totfager + (1 | id), family=binomial(link=logit), data=ctq)

> display(M1)
glmer(formula = Y ~ totfager + (1 | id), data = ctq, family = binomial(link = logit))
            coef.est  coef.se
(Intercept) -0.23     0.94
totfager    -0.44     0.15

Error terms:
 Groups   Name        Std.Dev.
 id       (Intercept) 3.68
 Residual             1.00
---
number of obs: 1688, groups: id, 266
AIC = 1336.6, DIC = 257.7
deviance = 794.2
```

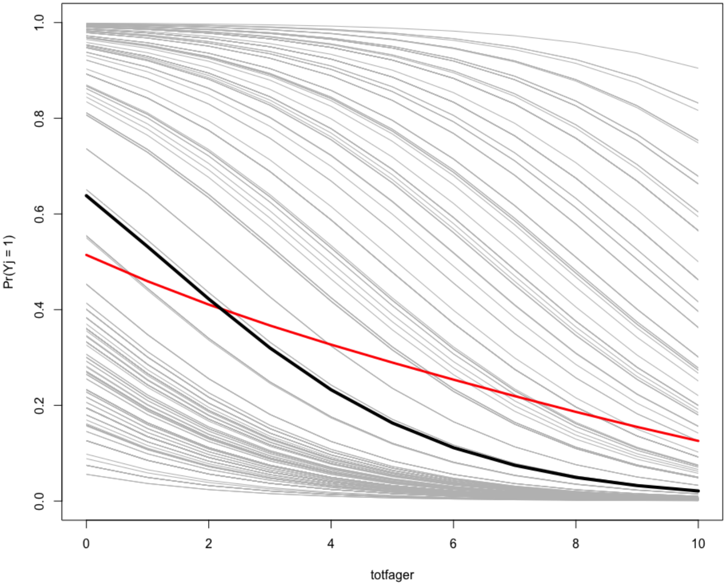
Fitting the model in R

```
> coef(M0)
$id
    (Intercept)   totfager
305   5.0808334 -0.4404643
309  -1.6677924 -0.4404643
311  -0.9908914 -0.4404643
313  -1.4195797 -0.4404643
314   1.8312968 -0.4404643
317  -0.8460499 -0.4404643
321  -1.9349872 -0.4404643
```

Plot of individual-level effect of X on logit scale



Probability scale



Population-averaged versus subject-specific effect of X

Subject-specific effect
Can write model like this

$$P(Y_{ij} = 1 \,|\, \alpha_i, X_i) \quad = \quad \frac{\exp(\alpha_i + \beta X_i)}{1 + \exp(\alpha_i + \beta X_i)}$$

For each individual i, the effect of X is β

Population-averaged versus subject-specific effect of X

Population-averaged effect

- To calculate PA effect, need to integrate (average) over distribution of α_i
- Recall that this distribution is $\mathcal{N}(\mu, \tau^2)$.
- Denote by $f(\alpha \,|\, \mu, \tau)$

The population-averaged effect is

$$P(Y_{ij} = 1 \,|\, X_i) \;=\; \int \frac{\exp(\alpha + \beta X_i)}{1 + \exp(\alpha + \beta X_i)}\, f(\alpha \,|\, \mu, \tau)\, d\alpha$$

Population-averaged versus subject-specific effect of X

When α has a normal distribution, can show that the integral is approximately

$$P(Y_{ij} = 1 \mid X_i) \approx \frac{\exp(\mu^* + \beta^* X_i)}{1 + \exp(\mu^* + \beta^* X_i)}$$

where

$$\beta^* = \frac{\beta}{(1 + .346\tau^2)^{1/2}}$$

In our example,

$$\begin{aligned} \widehat{\beta} &= -0.44 \\ \widehat{\beta}^* &= \frac{-0.44}{(1 + .346(3.68^2))^{1/2}} \approx -0.18 \end{aligned}$$

- At the individual level, nicotine dependence has a strong negative effect on probability of smoking cessation in a given week, associated with decreased log odds of quitting of $-.44$
- Averaged over the population, the effect is attenuated toward zero. The population-averaged effect represents a between-group comparison of those who differ by one unit in dependence score.

Model of treatment effect

- The structure of the model is motivated by the longitudinal data patterns
- Three parameters:
    - Overall intercept (quit rate prior to week 5)
    - Quit rate for control from weeks 5-12 (constant)
    - Difference in quit rate for treatment from weeks 5-12 (constant)

Model of treatment effect

Level 1

$$\begin{aligned} Y_{ij} &\sim \text{Ber}(\pi_{ij}) \\ \text{logit}(\pi_{ij}) &= \alpha_i + \beta T_j + \theta T_j Z_i \end{aligned}$$

Level 2

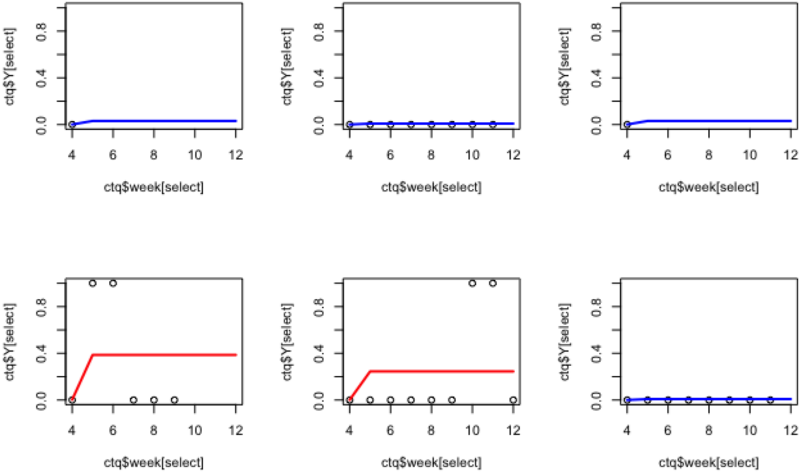
$$\alpha_i \sim \mathcal{N}(\mu, \tau^2)$$

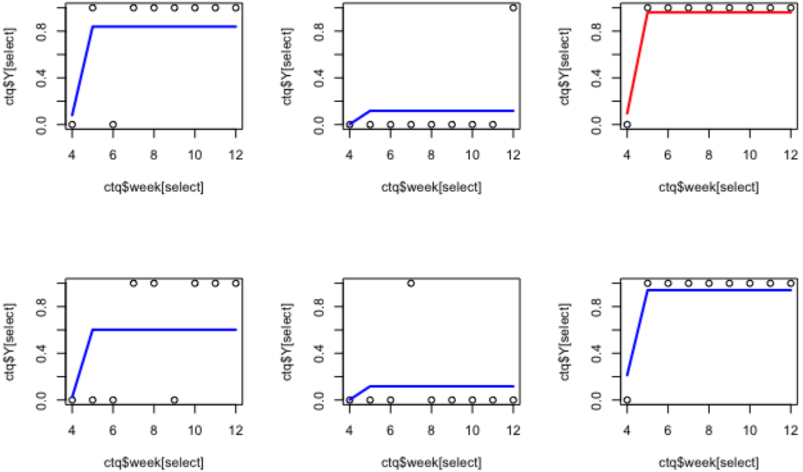
In this notation,

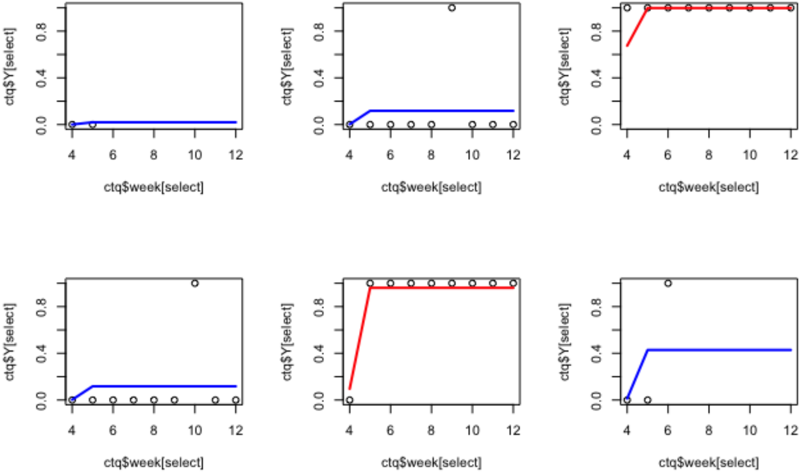
$$\begin{aligned} T_i &= 1 \text{ if week} \geq 5 \\ &= 0 \text{ if not} \end{aligned}$$

Fitting the model in R

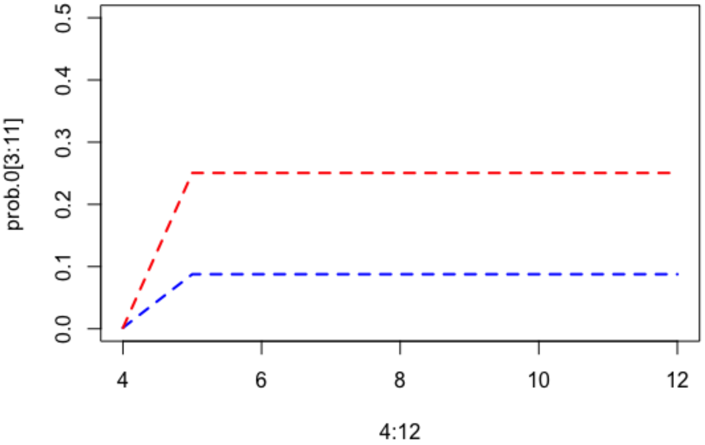
```
> M2 = glmer( Y ~ target_wk + target_wk:Z + (1  | id ), family=binomial(link=logit),

> display(M2)
glmer(formula = Y ~ target_wk + target_wk:Z + (1 | id), data = ctq,
    family = binomial(link = logit))
            coef.est coef.se
(Intercept) -7.52     0.80
target_wk    4.08     0.53
target_wk:Z  1.36     0.66

Error terms:
 Groups   Name        Std.Dev.
 id       (Intercept) 4.92
 Residual             1.00
---
number of obs: 1887, groups: id, 266
AIC = 1220.6, DIC = 84.9
deviance = 648.7
```


ctq$Y[select]
0.0
0.4
0.8
4
6
8
10
12
ctq$week[select]


ctq$Y[select]
0.0
0.4
0.8
4
6
8
10
12
ctq$week[select]


ctq$Y[select]
0.0
0.4
0.8
4
6
8
10
12
ctq$week[select]

Subject-specific effect



Population-averaged effect