Multilevel Logistic Regression for Binary Data

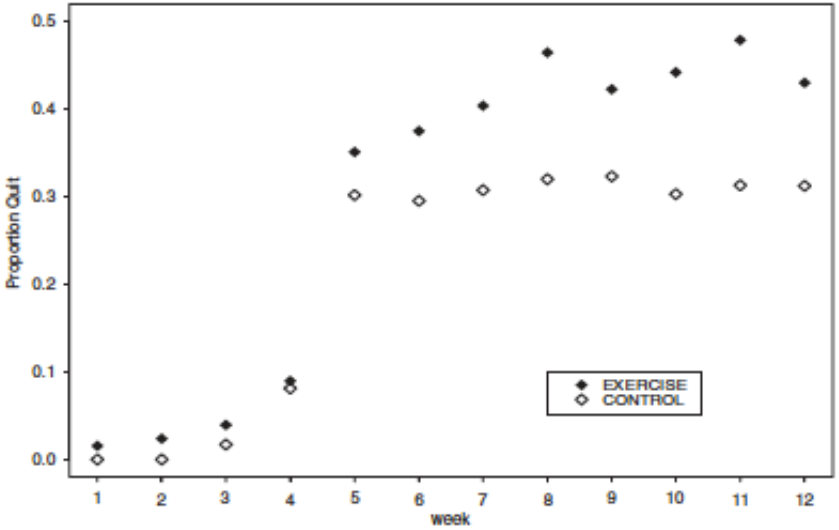
Nambari Short Course

16 July 2019

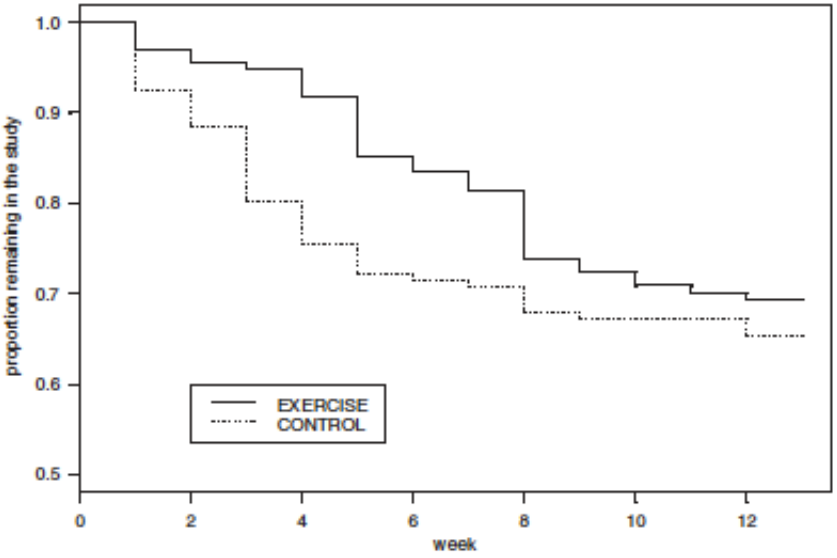
Motivating example: Smoking Cessation Study

- NIH-funded study to reduce smoking among sedentary women
- Roughly 300 individuals randomized to two arms:
  - Supervised exercise vs. Wellness education program
- Primary outcome
  - Weekly smoking status over 12 weeks
- Treatment comparison
  - Smoking rate at week 12 following baseline
- Analysis issues
  - Binary outcomes
  - Mean has some structure as a function of time
  - Large number of repeated measures

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 1: Intercept only

- Model structure follows very similarly to linear regression, except that we use a logit link.
- Each individual has their own intercept that summarizes $P(Y_{ij} = 1)$ across the j measurements.

Level 1

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

Level 2

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

Fitting the model in R

```
> M0 = glmer( Y ~ 1 + (1 | id), family=binomial, data=ctq)

> display(M0)
glmer(formula = Y ~ 1 + (1 | id), data = ctq, family = binomial)
coef.est  coef.se
   -3.23     0.50

Error terms:
 Groups   Name        Std.Dev.
 id       (Intercept) 4.02
 Residual             1.00
---
number of obs: 1887, groups: id, 266
AIC = 1433.9, DIC = 290.9
deviance = 860.4
```

Translating intercepts to subject-specific probabilities

- The coefficient α_i captures subject-specific propensity to quit smoking, but on the log odds scale.
- Can translate to the probability scale with inverse logit function

```
> alpha.hat = coef(M0)$id
> prob.hat  = exp(alpha.hat) / (1 + exp(alpha.hat))
>
> cbind(alpha.hat, prob.hat)
     (Intercept) (Intercept)
305    1.7704047 0.854507998
309   -4.6324170 0.009637427
311   -4.6324170 0.009637427
313   -2.1564835 0.103726914
314   -2.1564835 0.103726914
317   -3.6428610 0.025509570
321   -4.6324170 0.009637427
324   -4.6324170 0.009637427
```

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

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

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)

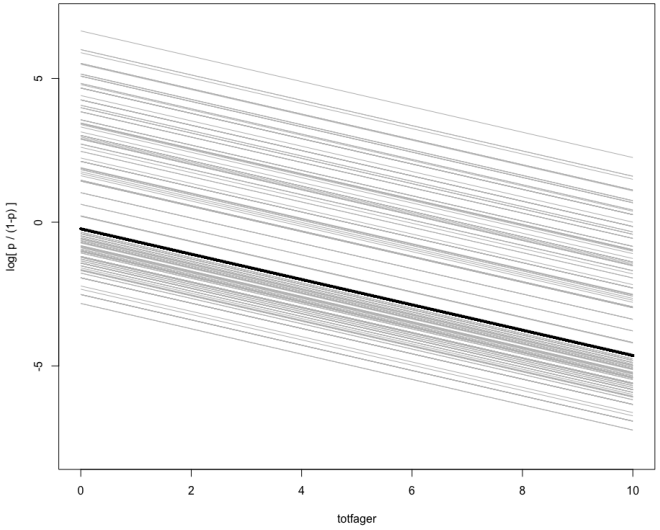
Fitting the model in R

```
> 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