Lab 3 Mini-Lecture

Categorical MLM

1

Today's Roadmap

- General feedback and review
  - Assignment 2
    - Review models and excel file
- Exercise 3
  - Starting today
  - Will finish on Monday

2

Thailand Education Data

- 7516 pupils in primary schools
  - Example data from HLM
- Outcome variable dichotomous
  - Repeat a class (1) or not (0)
- Pupil level predictors
  - sex (1=male, 0=female)
  - pre-primary education (PPED: 1=yes, 0=nc
- School level predictor
  - Mean SES

3

Generalized linear model

- To analyze dichotomous multilevel data in R, we need to use the function glmer() instead of lmer()
  - where 'g' stands for generalized of the generalized linear model.
- Regression portion similar to lmer
  - Follows the same structure we have been using already

4

glmer

- Need to specify appropriate link function
- Also, need to adjust the number of points per axis for evaluating the adaptive Gauss-Hermite approximation to the log-likelihood using nAGQ = 10 in order to ensure convergence of the model

- https://www.rdocumentation.org/packages/lme4/versions/1.1-26/topics/glmer

5

Let's get started

- Read in and view the data
- Specify the model with the random intercept and no predictors.
  - Compare to table from class

```
M1 <- glmer(rep1~1 + (1|schoolid), data = dat, family = binomial, nAGQ = 10)
```

6

Thailand Education Data

Model	Empty model		Predictors level 1, fixed effects		
Predictor	par est	SE	par est	SE	Odds ratio
Fixed part					
threshold	-2.23	0.08	-2.23	0.10	
Sex (female=ref)			0.54	0.07	1.71
PPED (no=ref)			-0.64	0.10	0.53
Mean SES					
Random part		SD		SD	
Intercept school level	1.73	1,31	1.69	1,30	
Model fit					
Deviance	5537.4		5443.5		
AIC	5541.4		5451.5		

7

Continue with exercise

- Set up the model with the level 1 predictors male and pped.
- Once parameters match previous table, obtain the variance of the linear predictor in either R or SPSS using the obtained regression coefficients for the level 1 predictors (and intercept).
- Variance of linear predictor of fixed part σ_F^2 can be calculated in R or SPSS using estimates of regression coefficients

8

Recall example from class: Variance of the linear predictor

ID	sex	age	…	$\gamma_{00} + \gamma_{10}\boldsymbol{sex}_{ij} + \gamma_{10}\boldsymbol{age}_{ij}$
1	0	23		-2.8
2	0	20		-2.71
3	1	19		-2.27
4	1	25		-2.45
5	1	24		-2.42
6	0	28		-2.95
7	1	22		-2.36

$\sigma_F^2 =$ <- var

$$Y_{ij} = \gamma_{00} + \gamma_{10}\boldsymbol{sex}_{ij} + \gamma_{20}\boldsymbol{age}_{ij}$$

$$Y_{ij} = -2.11 + 0.41\ \boldsymbol{sex}_{ij} - 0.03\ \boldsymbol{age}_{ij}$$

9

Helpful reminder of definitions of logit and odds from class slides

p_{ij}=probability of repeating a class

$$\text{logit}(p_{ij}) = -2.237 + 0.536 sex_{ij} - 0.642 pped_{ij} + u_{0j}$$

$$odds = \frac{p_{ij}}{1 - p_{ij}}$$

$$odds\ ratio = \frac{odds_{males}}{odds_{females}} = e^{0.536} = 1.709$$

10

Interpretation of model parameters

Qualitative interpretation for sex

- Males have higher probability of repeating a class than females

Quantitative interpretation for sex

- On logit scale:
  - Logit probability of repeating a class is 0.536 higher for males than for females (effect is additive on logit scale)
- Using odds ratios
  - Odds of repeating a class are 1.709 times as high for males as for females (effect is multiplicative on odds scale)

In the exam, you put both interpretation. The general qualitative, & the quantitative w/ all the stats, ie odds ratio, significance value... etc. In scoring sheet, there needs to be both otherwise deduction, highly recommend, in the exam, you interpret in terms of odds ratio

Thailand Education Data

- $\rho = \frac{\sigma^2_{u0}}{\sigma^2_{u0}+\sigma^2_R} = \frac{1.725}{1.725+3.29} = 0.331$

- Estimates of variance components
  - $\sigma^2_{u0} = 1.627$
  - $\sigma^2_R = 3.290$
  - $\sigma^2_F = 0.176$
- Explained and unexplained variance
  - Explained: $\frac{\sigma^2_F}{(\sigma^2_F+\sigma^2_{u0}+\sigma^2_R)} = 0.035$
  - Unexplained level 1: $\frac{\sigma^2_R}{(\sigma^2_F+\sigma^2_{u0}+\sigma^2_R)} = 0.646$
  - Unexplained level 2: $\frac{\sigma^2_{u0}}{(\sigma^2_F+\sigma^2_{u0}+\sigma^2_R)} = 0.319$