GLMMs Summary

← "conditional"

Wednesday, November 22, 2023 16:12

Take GLMs to the next level by introducing random intercepts.

Random Intercept Logistic Regression Model

$y_{ij} \sim \text{Binomial}(p_{ij})$

$p_{ij} = E[y_{ij} \mid \theta_i]$

$i = 1, \ldots, n$ group
$j = 1, \ldots, r_i$ obs. within group

$\text{logit}(p_{ij}) = \beta_0 + \theta_i + x_{ij}'\beta$

NO ERROR TERM

- β_0: baseline log odds overall
- x_{ij} = $p \times 1$ vector of covariates
- β: vector of regression coefficients

$\theta_i \overset{iid}{\sim} N(0, \tau^2)$ → variation of baseline log odds between groups

The difference between group-specific baseline log odds and β_0

Likelihood:

y is data, → all y_i independent

θ is vector of ALL RANDOM EFFECTS, $[y, \theta]$ is joint density

$$[y, \theta] = \prod_{i=1}^{n} [y_i, \theta_i]$$ → meaning joint density of $[y_{i1}, \ldots, y_{ir_i}, \theta_i]$ (r_i: # obs for subject i)

$$= \prod_{i=1}^{n} [y_i \mid \theta_i][\theta_i]$$

$$= \prod_{i=1}^{n} \left(\prod_{j=1}^{r_i} [y_{ij} \mid \theta_i] \right) [\theta_i]$$ (y_{ij} conditionally independent given random effects)

$$= \prod_{i=1}^{n} \left(\prod_{j=1}^{r_i} [y_{ij} \mid \theta_i] \right) (2\pi\tau^2)^{-1/2} \exp\left(\frac{1}{-2\tau^2} \theta_i^2 \right)$$

$$L(\beta, \beta_0, \tau^2 \mid y) = \prod_{i=1}^{n} \int \prod_{j=1}^{r_i} [y_{ij} \mid \theta_i] * (2\pi\tau^2)^{-1/2} \exp\left(\frac{1}{2\tau^2} \theta_i^2 \right) d\theta_i$$

$$\prod_{j=1}^{r_i} [y_{ij} \mid \theta_i] = \prod_{j=1}^{r_i} p_{ij}^{y_{ij}} * (1 - p_{ij})^{1 - y_{ij}}$$

$$= \prod_{j=1}^{r_i} \left(\frac{e^{\beta_0 + \theta_i + x_{ij}'\beta}}{1 + e^{\beta_0 + \theta_i + x_{ij}'\beta}} \right)^{y_{ij}} \times \left(\frac{1}{1 + e^{\beta_0 + \theta_i + x_{ij}'\beta}} \right)^{1 - y_{ij}}$$

$$= \prod_{i=1}^{n} \int \prod_{j=1}^{r_i} \left(\frac{e^{\beta_0 + \theta_i + x_{ij}'\beta}}{1 + e^{\beta_0 + \theta_i + x_{ij}'\beta}} \right)^{y_{ij}} * \left(\frac{1}{1 + e^{\beta_0 + \theta_i + x_{ij}'\beta}} \right)^{1 - y_{ij}} * (2\pi\tau^2)^{-1/2} \exp\left(-\frac{1}{2\tau^2} \theta_i^2 \right) d\theta_i$$

→ ~~For~~ GLMM, convergence becomes an issue here.

→ increase nAGQ in R to help it converge?

→ In general, slopes not comparable

Baseline log odds across subjects: Intercept estimate $\pm 1.96 (\sqrt{\tau^2})$

Random Intercept Poisson Model

outcome e.g. death count ← $y_{stk} \sim \text{Poisson}(\lambda_{stk})$

$\theta_s \overset{iid}{\sim} N(0, \tau^2)$ ↳ variation between groups

θ_s: group-specific deviation in baseline log expected [outcome]

$$\log(E[y_{stk} \mid \theta_s]) = \log \lambda_{stk} = \beta_0 + \theta_s + \beta_1 I[sex_k = F] + \beta_2 I[race_k = W] + \beta_3 I[sex_k = F] I[race_k = W] + \beta_4 year_t$$

- β_0: log expected count at baseline for the avg county (non-white males in 1979)
- e^{β_0} will be baseline expected count.
- ↳ e^{β_1} = ratio of deaths for non-white F to non-white M in model accounting for random intercepts, ..., ...
- ↳ e^{β_3} = relative rate modification in deaths for white F
- ↳ e^{β_4} = rate ratio of ... for a 1-yr increase in a model controlling for ...

CI for baseline counts: $e^{\text{Intercept estimate} \pm 1.96(\tau \text{ aka random effects std. dev})}$

$e^{\text{cont. var estimate}}$ → __ rate was increasing by this per [year]

or $100 * (e^{\text{cont. var estimate}} - 1)$ = __ % per [year]

$e^{\beta_0 + \beta_t * \ldots}$ = expected __ count for ____ in 1980...

GOF test: H_0: Model adequately fits the data

H_1: Lack of fit

Random Intercept Poisson Model WITH POPULATION OFFSET

← Same model but add $\log pop_{stk}$ which is an offset variable

$$\lambda_{stk} = e^{\log pop_{stk} + \beta_0 + \theta_s + \ldots}$$

$$= pop_{stk} * e^{\beta_0 + \theta_s + \ldots}$$

Then,

$$\frac{\lambda_{stk}}{pop_{stk}} = e^{\beta_0 + \theta_s + \ldots}$$

e^{β_0} is per capita deaths instead of expected counts

↳ $e^{\beta_0} = \lambda_i / pop_i$

Quick derivation:

$$\log \lambda_i = \beta_0 + \log pop_i$$

$$\log \lambda_i - \log pop_i = \beta_0$$

$$\log\left(\frac{\lambda_i}{pop_i}\right) = \beta_0$$

$$e^{\beta_0} = \frac{\lambda_i}{pop_i}$$

3.3 What is the odds ratio and 95% confidence interval for treatment=Intervention in age group 25-34 versus Control in age group 25-34?

1.72 (0.378, 7.78)

```
control <- 2.0624
intervention <- -1.5228

q3_3 <- exp(2.0624 + -1.5228)
q3_3
```

```
## [1] 1.72
```

```
q3_3var <- vcov(q3_2)
q3_3var2 <- q3_3var["treatmentIntervention", "treatmentIntervention"] +
            q3_3var["treatmentIntervention:agegroup25-34", "treatmentIntervention:agegroup25-34"] +
            2 * q3_3var["treatmentIntervention", "treatmentIntervention:agegroup25-34"]

q3_3 * exp(- 1.96 * sqrt(q3_3var2))
```

```
## [1] 0.378
```

```
q3_3 * exp(1.96 * sqrt(q3_3var2))
```

```
## [1] 7.78
```