

Wednesday, November 22, 2023 19:12



## Model Fitting and Interpretations

We specify the multi level model with two random intercept components. (Here, child ID is unique, so the below code is equivalent to `(1|childid)` (1|childid school, we fit only

```

i <- fix ~ 1| school + year + sex + school * year + sex * school +
  (1|school) + (1|childid, new = 0)

```

# summary (fit)

Random effects		
Group	Mean	Variance (SD)
Intercept	0.0000000	0.0000000 (0.0000000)
Intercept	0.0000000	0.0000000 (0.0000000)
Residual		0.0000000 (0.0000000)

Note that the random effects of *child* and *school* are assumed to be independent.

- Within child correlation =  $\frac{0.0000000}{0.0000000 + 0.0000000} = 0.00$
- Within school correlation =  $\frac{0.0000000}{0.0000000 + 0.0000000} = 0.00$

[illegible]

### Three-Level Logistic Random Intercept Model

Intergenic – Binomial [1]

$$\text{logit}(\mu_{ijk}) = \beta_0 + \alpha_i + \eta_{ij} + \beta_1 X_{ijk} + \beta_2 X_{ijk}^2 + \beta_3 X_{ijk}^3$$

$\mu_{ijk} = \text{mean}$   
 $\eta_{ij} \sim N(0, \sigma^2)$   
 $\alpha_i \sim N(0, \sigma^2)$   
 $\sigma^2 = \text{variance}$   
 $\alpha_i = \text{intercept}$

- $\mu_{ijk} = E(Y_{ijk})$
- $\text{Intercept} = \beta_0 + \alpha_i + \eta_{ij}$
- $\alpha_i = \text{family-level intercept}$
- $\eta_{ij} = \text{individual-level intercept}$
- $\sigma^2 = \text{variance}$
- $\sigma^2 = \text{between-family variation in baseline log odds}$
- $\sigma^2 = \text{between-individual variation in baseline log odds}$
- Only two normal random effects:  $\alpha_i \sim N(0, \sigma^2)$ ,  $\eta_{ij} \sim N(0, \sigma^2)$
- $\sigma^2 = \text{variance}$
- We can think of  $\alpha_i$  and  $\eta_{ij}$  as parameters that control for **unmeasured confounders** at the community- and family-level. One structural model is another's confounder structure!

[illegible]

**Point estimates**

	Estimate	95% CI	OR (95% CI)
Constant	-1.7624	-2.8613 - 0.6535	0.0000
Age	0.0000	-0.0000 - 0.0000	1.0000
Gender	0.2363	-0.1471 - 0.6099	0.0000
Married	0.0000	-0.0000 - 0.0000	1.0000
Education	0.0000	-0.0000 - 0.0000	1.0000
Health	0.2772	-0.0487 - 0.6008	0.0000
Income	0.0000	-0.0000 - 0.0000	1.0000
Year	-0.0000	-0.0000 - 0.0000	0.0000
Year <sup>2</sup>	-0.0000	-0.0000 - 0.0000	0.0000

After controlling for within-family and within-community correlation,

- odds of immigration decreased in rural communities (OR:  $e^{-0.67} = 0.51$ , 24% decrease), and communities with higher percentage of indigenous population.  $1 - e^{(-0.0001 \times 10000)} = 0.933$ .
- Higher immigration rate was associated with families where the mother (OR:  $e^{0.28} = 1.3$ ) or the husband (OR:  $e^{0.28} = 1.4$ ) were primary school graduates.
- Significant effects for secondary schools (smaller sample size).
- Children born during the campaign had a higher immigration rate (OR:  $e^{1.26} = 3.50$ ).

What is the odds ratio for a child at least 2 years old ( $\text{kid2p}=1$ ) versus less than 2 years old ( $\text{kid2p}=0$ ) for a child from the same family and community, holding other variables constant?

$$e^{1.2815+\alpha_i+\alpha_{iU}+\beta'x_{iU}}/e^{\beta+\alpha_i+\alpha_{iU}+\beta'x_{iU}} = e^{1.2815}$$

$$e^{1.2815} = 3.60, \text{ 95\% CI: } e^{1.2815 \pm 1.96 \times 0.1581} = [2.642, 4$$

What is the odds ratio and 95% CI in immunization between two children from the same community and same mother with:

- child B = over 2 years old at interview, mother's husband had primary education  $661p = 1$
- child A = under 2 years old at interview, mother's husband had no primary education

$$\text{logit}(p_B) - \text{logit}(p_A) = 1.2815 + 0.3771$$

$$\text{OR for child B versus child A} = e^{1.2815+0.3771} = 5.25$$

[illegible]

3. Fit of a statistical model that uses logistic regression with response `death_pathogen_juvenile`, with random effects for `year`, `sex`, `pathogen`, `stage`, and their interaction, and including `death`, `ln_population`, `shrub_leaf_area`, `all_type`, `color_juvenile`, and `fallout_paper` as fixed effects. Use  $\beta_{00}$ ,  $\beta_{10}$ , and  $\beta_{20}$  denote the random effects, which we assume are independent for factors, use the 1 standard deviation level and include a separate coefficient for each level.

$$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = (\beta_0 + \beta_1 \text{Depth}(\text{Middle})_{ij} + \beta_2 \text{Depth}(\text{Top})_{ij} + \beta_3 \text{RR\_population}_{ij} \\ + \beta_4 \text{shared\_letter}_{ij} + \beta_5 \text{all\_type}_{ij} + \beta_6 \text{water\_source\_know}_{ij} + \beta_7 \text{water\_source\_public}_{ij} + \beta_8 \text{toilet\_paper}_{ij} \\ + \beta_9 + \beta_{10} + \beta_{11})$$

$\mathbb{P}_{i,j}(k)$  is the probability of `depth_pathogen_presence` to take  $i$ , `pathogen_target`  $j$  and depth  $k$ .

$\mathbb{E}_P \sup_{f \in \mathcal{F}} \sum_{t=1}^T \ell_t(f) = O(\sqrt{T})$  for all  $P \in \mathcal{P}$  and  $\mathcal{F}$  is a function class.

$\hat{\mu}_{00}$  is the lesion-specific diameter in baseline expected log odds. It is fit with  $N(\hat{\mu}_{00}, \sigma^2)$  with  $\sigma^2$  being variance between subjects.

$\sigma_{ij}$  is the pathogen target specific variation in baseline expected log odds. It is  $\text{IG}(\tau^2)$  with  $N(\theta, \tau^2)$  with  $\tau^2$  being variance between pathogen targets.

$\beta_{\text{path}}$  is the pathogen target and host–microbe specific deviation in baseline expected log odds. It is GD with  $N(0, \sigma^2)$  where  $\sigma^2$  is variance between the specific combinations of pathogen targets and tissues.

$\beta_5$  is the log odds ratio associated with middle depth while controlling for other covariates.  $\beta_6$  is the log odds ratio associated with log depth while

controlling for other covariates.  $\beta_{10}$  is the log odds ratio associated with WHI appearance while controlling for other covariates.  $\beta_{11}$  is the log odds ratio associated with shared\_name while controlling for other covariates.  $\beta_{12}$  is the log odds ratio associated with pit\_type while controlling for other covariates.  $\beta_{13}$  is the log odds ratio associated with water\_source while controlling for other covariates.  $\beta_{14}$  is the log odds ratio associated with

$\beta_1$  is the log odds ratio associated with water\_purifier, while controlling for other covariates.  $\beta_2$  is the log odds ratio associated with water\_source\_public while controlling for other covariates.  $\beta_3$  is the log odds ratio associated with toilet\_paper while controlling for other covariates.

$$A_2 = I_2 - A_1$$

After accounting for correlations among observations within the same labline, the same pathogen target, and the same combination of labline and pathogen target.

- 1. Example:
  - Null hypothesis: the disease probability of the presence of pathogen is 0.5
  - Conclusion: Since  $\alpha = 0.0101 < 0.05$ , we reject the null hypothesis and conclude that the baseline probability of the presence of pathogen is significantly different from 0.5
- 2.  $H_0$  hypothesis:
  - Null hypothesis: the coefficient of  $\beta_{pathogen}$  is 0.0. *Null hypothesis does not significantly influence the presence of pathogen*
  - Conclusion: Since  $\alpha = 0.0010 < 0.05$ , we reject the null hypothesis and conclude that the coefficient of  $\beta_{pathogen}$  significantly differs from zero. *Cr.  $\beta_{pathogen}$  significantly influences the presence of pathogen*
- 3.  $\beta_{pathogen}$ 
  - Null hypothesis: the presence of pathogen in a dataset with a disease source does not significantly differ from that in a dataset with a compound water source
  - Conclusion: Since  $\alpha = 0.0014 < 0.05$ , we reject the null hypothesis and conclude that the probability of the presence of pathogen in a dataset with a disease source is significantly different from that in a dataset with a compound water source

9. What is the probability of pathogen presence in a household with 10 people with a public water source, all other variables set to the baseline?

$$\text{area} = -2.82612 + (0.17987 \times 10) + 0.7282$$

$$\exp(\eta_{it}) / (1 + \exp(\eta_{it}))$$