Agenda

1. More on multilevel R formulas
2. Multilevel generalized linear models (MLGLMs)
3. MLGLMs in R

More on multilevel R formulas

Building a two-level model

A simple two-level model

$$S_{ik} \sim \text{Norm}(\mu_{ik}, \sigma)$$

$$\mu_{ik} = \beta_{0k} + \beta_1 Age_i + \beta_2 Female_i + \beta_{3k} Black_i$$

$$\beta_{0k} = \gamma_{00} + \gamma_{01} TeacherBlack_k + \gamma_{02} PropBlack_k + \eta_{0k}$$

$$\beta_{3k} = \gamma_{30} + \gamma_{31} TeacherBlack_k + \gamma_{32} PropBlack_k + \eta_{3k}$$

Expanded notation

$$\mu_{ik} = \boxed{\beta_{0k}} + \beta_1 Age_i + \beta_2 Female_i + \boxed{\beta_{3k}} Black_i$$

$$\boxed{\beta_{0k}} = \gamma_{00} + \gamma_{11} TeacherBlack_k + \gamma_{051} PropBlack_k + \eta_{0k}$$

$$\boxed{\beta_{3k}} = \gamma_{30} + \gamma_{31} TeacherBlack_k + \gamma_{32} PropBlack_k + \eta_{3k}$$

$$\begin{aligned} \mu_{ik} = & \left(\gamma_{00} + \gamma_{01} TeacherBlack_k + \gamma_{02} PropBlack_k + \eta_{0k}\right) \\ & + \beta_1 Age_i \\ & + \beta_2 Female_i \\ & + \left(\gamma_{30} + \gamma_{31} TeacherBlack_k + \gamma_{32} PropBlack_k + \eta_{0k}\right) Black_i \end{aligned}$$

Expanded notation

$$
\begin{aligned}
\mu_{ik} = {} & (\gamma_{00} + \gamma_{01} TeacherBlack_k + \gamma_{02} PropBlack_k + \eta_{0k}) \\
& + \beta_1 Age_i \\
& + \beta_2 Female_i \\
& + (\gamma_{30} + \gamma_{31} TeacherBlack_k + \gamma_{32} PropBlack_k + \eta_{0k}) Black_i
\end{aligned}
$$

$$
\begin{aligned}
s_{ik} = {} & \gamma_{00} + \gamma_{01} TeacherBlack_k + \gamma_{02} PropBlack_k \\
& + \beta_1 Age_i + \beta_2 Female_i + \gamma_{30} Black_i \\
& + \gamma_{31} TeacherBlack_k Black_i + \gamma_{32} PropBlack_k Black_i \\
& + \eta_{0k} + \eta_{3k} Black_i + \varepsilon_i
\end{aligned}
$$

Expanded notation

Fixed effects

Explained variation in outcome variable.

Describes the way that outcome and predictor variables co-vary.

$$s_{ik} = \gamma_{00} + \gamma_{01} TeacherBlack_k + \gamma_{02} PropBlack_k$$
$$+ \beta_1 Age_i + \beta_2 Female_i + \gamma_{30} Black_i$$
$$+ \gamma_{31} TeacherBlack_k Black_i + \gamma_{32} PropBlack_k Black_i$$
$$+ \eta_{0k} + \eta_{3k} Black_i + \varepsilon_i$$

Random effects

Unexplained variation in outcome variable.

Described in terms of individual variability and different types of group variability.

Building an R formula

With level-two covariates,
interactions need to be specified

```
student_reading_score ~
        student_age_s + student_female
        student_re_black*teacher_re_black +
        student_re_black*class_prop_black +
        (1 + student_re_black | teacher_id)
```

$$
\begin{aligned}
s_{ik} = \gamma_{00} &+ \gamma_{01} TeacherBlack_k + \gamma_{02} PropBlack_k \\
&+ \beta_1 Age_i + \beta_2 Female_i + \gamma_{30} Black_i \\
&+ \gamma_{31} TeacherBlack_k Black_i + \gamma_{32} PropBlack_k Black_i \\
&+ \eta_{0k} + \eta_{3k} Black_i + \varepsilon_i
\end{aligned}
$$

Building an R formula

Outcome variable

```
student_reading_score ~
        student_age_s + student_female
        student_re_black*teacher_re_black +
        student_re_black*class_prop_black +
        (1 + student_re_black | teacher_id)
```

$$
\begin{aligned}
S_{ik} = {} & \gamma_{00} + \gamma_{01} TeacherBlack_k + \gamma_{02} PropBlack_k \\
& + \beta_1 Age_i + \beta_2 Female_i + \gamma_{30} Black_i \\
& + \gamma_{31} TeacherBlack_k Black_i + \gamma_{32} PropBlack_k Black_i \\
& + \eta_{0k} + \eta_{3k} Black_i + \varepsilon_i
\end{aligned}
$$

Building an R formula

Global intercept (included automatically)

```
student_reading_score ~
        student_age_s + student_female
        student_re_black*teacher_re_black +
        student_re_black*class_prop_black +
        (1 + student_re_black | teacher_id)
```

$$
\begin{aligned}
s_{ik} = \boxed{\gamma_{00}} &+ \gamma_{01} TeacherBlack_k + \gamma_{02} PropBlack_k \\
&+ \beta_1 Age_i + \beta_2 Female_i + \gamma_{30} Black_i \\
&+ \gamma_{31} TeacherBlack_k Black_i + \gamma_{32} PropBlack_k Black_i \\
&+ \eta_{0k} + \eta_{3k} Black_i + \varepsilon_i
\end{aligned}
$$

Building an R formula

Interactions (*) automatically include standalone terms

```
student_reading_score ~
          student_age_s + student_female
          student_re_black*teacher_re_black +
          student_re_black*class_prop_black +
          (1 + student_re_black | teacher_id)
```

$$s_{ik} = \gamma_{00} + \gamma_{01} TeacherBlack_k + \gamma_{02} PropBlack_k$$
$$+ \beta_1 Age_i + \beta_2 Female_i + \gamma_{30} Black_i$$
$$+ \gamma_{31} TeacherBlack_k Black_i + \gamma_{32} PropBlack_k Black_i$$
$$+ \eta_{0k} + \eta_{3k} Black_i + \varepsilon_i$$

Building an R formula

Interactions (*) automatically include standalone terms

```
student_reading_score ~
        student_age_s + student_female
        student_re_black*teacher_re_black +
        student_re_black*class_prop_black +
        (1 + student_re_black | teacher_id)
```

$$
\begin{aligned}
s_{ik} = &\ \gamma_{00} + \gamma_{01} TeacherBlack_k + \boxed{\gamma_{02} PropBlack_k} \\
&+ \beta_1 Age_i + \beta_2 Female_i + \boxed{\gamma_{30} Black_i} \\
&+ \gamma_{31} TeacherBlack_k Black_i + \boxed{\gamma_{32} PropBlack_k Black_i} \\
&+ \eta_{0k} + \eta_{3k} Black_i + \varepsilon_i
\end{aligned}
$$

Building an R formula

Redundant terms (`student_re_black`) are not added twice

```
student_reading_score ~
        student_age_s + student_female
        student_re_black*teacher_re_black +
        student_re_black*class_prop_black +
        (1 + student_re_black | teacher_id)
```

$$
\begin{aligned}
s_{ik} = {} & \gamma_{00} + \gamma_{01} TeacherBlack_k + \gamma_{02} PropBlack_k \\
& + \beta_1 Age_i + \beta_2 Female_i + \gamma_{30} Black_i \\
& + \gamma_{31} TeacherBlack_k Black_i + \gamma_{32} PropBlack_k Black_i \\
& + \eta_{0k} + \eta_{3k} Black_i + \varepsilon_i
\end{aligned}
$$

Building an R formula

Random effects use pipe notation (|)

```
student_reading_score ~
        student_age_s + student_female
        student_re_black*teacher_re_black +
        student_re_black*class_prop_black +
        (1 + student_re_black | teacher_id)
```

$$
\begin{aligned}
s_{ik} = {} & \gamma_{00} + \gamma_{01} TeacherBlack_k + \gamma_{02} PropBlack_k \\
& + \beta_1 Age_i + \beta_2 Female_i + \gamma_{30} Black_i \\
& + \gamma_{31} TeacherBlack_k Black_i + \gamma_{32} PropBlack_k Black_i \\
& + \boxed{\eta_{0k} + \eta_{3k} Black_i + \varepsilon_i}
\end{aligned}
$$

Building an R formula

Grouping elements after the pipe

```
student_reading_score ~
        student_age_s + student_female
        student_re_black*teacher_re_black +
        student_re_black*class_prop_black +
        (1 + student_re_black | teacher_id)
```

$$
\begin{aligned}
s_{ik} = \gamma_{00} &+ \gamma_{01} TeacherBlack_k + \gamma_{02} PropBlack_k \\
&+ \beta_1 Age_i + \beta_2 Female_i + \gamma_{30} Black_i \\
&+ \gamma_{31} TeacherBlack_k Black_i + \gamma_{32} PropBlack_k Black_i \\
&+ \eta_{0k} + \eta_{3k} Black_i + \varepsilon_i
\end{aligned}
$$

Building an R formula

Random intercepts indicated with constant (1)

```
student_reading_score ~
        student_age_s + student_female
        student_re_black*teacher_re_black +
        student_re_black*class_prop_black +
        (1 + student_re_black | teacher_id)
```

$$
\begin{aligned}
s_{ik} &= \gamma_{00} + \gamma_{01} TeacherBlack_k + \gamma_{02} PropBlack_k \\
&+ \beta_1 Age_i + \beta_2 Female_i + \gamma_{30} Black_i \\
&+ \gamma_{31} TeacherBlack_k Black_i + \gamma_{32} PropBlack_k Black_i \\
&+ \boxed{\eta_{0k}} + \eta_{3k} Black_i + \varepsilon_i
\end{aligned}
$$

Building an R formula

Random-slope variables included in grouping expression

```
student_reading_score ~
        student_age_s + student_female
        student_re_black*teacher_re_black +
        student_re_black*class_prop_black +
        (1 + student_re_black | teacher_id)
```

$$
\begin{aligned}
s_{ik} = & \gamma_{00} + \gamma_{01} TeacherBlack_k + \gamma_{02} PropBlack_k \\
& + \beta_1 Age_i + \beta_2 Female_i + \gamma_{30} Black_i \\
& + \gamma_{31} TeacherBlack_k Black_i + \gamma_{32} PropBlack_k Black_i \\
& + \eta_{0k} + \boxed{\eta_{3k} Black_i} + \varepsilon_i
\end{aligned}
$$

Generalized multilevel models

Generalized multilevel models

Generalized multilevel linear models

Simply add a link function and change the outcome distribution.

E.g. modeling whether a student did better on the math test than the reading test (M_{ik}).

$$M_{ik} \sim \text{Binomial}(1, p_{ik})$$
$$\text{logit}(p_{ik}) = \beta_{0k} + \beta_{1k} Age_i + \beta_{2k} Female_i$$

$$\beta_{0k} = \gamma_{00} + \gamma_{01} TExp_k + \gamma_{02} TFemale_k + \eta_{0k}$$
$$\beta_{1k} = \gamma_{10} + \eta_{1k}$$
$$\beta_{2k} = \gamma_{20} + \gamma_{21} TFemale_k + \eta_{2k}$$

Generalized multilevel models

This is where the 'generalized' part of the model lives

$$M_{ik} \sim \text{Binomial}(1, p_{ik})$$

$$\text{logit}(p_{ik}) = \beta_{0k} + \beta_{1k}Age_i + \beta_{2k}Female_i$$

$$\beta_{0k} = \gamma_{00} + \gamma_{01}TExp_k + \gamma_{02}TFemale_k + \eta_{0k}$$

$$\beta_{1k} = \gamma_{10} + \eta_{1k}$$

$$\beta_{2k} = \gamma_{20} + \gamma_{21}TFemale_k + \eta_{2k}$$

Generalized multilevel models

$$M_{ik} \sim \text{Binomial}(1, p_{ik})$$

This is the same as a standard (Gaussian) multilevel model

$$\text{logit}(p_{ik}) = \beta_{0k} + \beta_{1k} Age_i + \beta_{2k} Female_i$$

$$\beta_{0k} = \gamma_{00} + \gamma_{01} TExp_k + \gamma_{02} TFemale_k + \eta_{0k}$$

$$\beta_{1k} = \gamma_{10} + \eta_{1k}$$

$$\beta_{2k} = \gamma_{20} + \gamma_{21} TFemale_k + \eta_{2k}$$

Generalized multilevel models

Generalized multilevel linear models

Simply add a link function and change the outcome distribution.

E.g. modeling whether a student did better on the math test than the reading test (M_{ik}).

$$M_{ik} \sim \text{Binomial}(1, p_{ik})$$

$$\text{logit}(p_{ik}) = \beta_{0k} + \beta_{1k} Age_i + \beta_{2k} Female_i$$

$$\beta_{0k} = \gamma_{00} + \gamma_{01} TExp_k + \gamma_{02} TFemale_k + \eta_{0k}$$

$$\beta_{1k} = \gamma_{10} + \eta_{1k}$$

$$\beta_{2k} = \gamma_{20} + \gamma_{21} TFemale_k + \eta_{2k}$$

All coefficients and parameters are affected by link function

Interpretation requires careful thinking about the ways that coefficients affect p_{ik}.

Interpreting *direction* of effect (positive vs. negative) is still straightforward. E.g. a strong positive estimate on γ_{02} would suggest that female teachers do a relatively better job of teaching mathematics.

Generalized multilevel models

Gamma–Poisson `family = negbinomial`

$$Y_{ik} \sim \text{Pois}(\lambda_{ik})$$
$$\lambda_{ik} \sim \text{Gamma}(\mu_{ik}, \theta)$$
$$\log(\mu_{ik}) = \beta_{0k} + \beta_{1k} Age_i + \beta_{2k} Female_i$$

$$\beta_{0k} = \gamma_{00} + \gamma_{01} TExp_k + \gamma_{02} TFemale_k + \eta_{0k}$$
$$\beta_{1k} = \gamma_{10} + \eta_{1k}$$
$$\beta_{2k} = \gamma_{20} + \gamma_{21} TFemale_k + \eta_{2k}$$

Multinomial `family = categorical`

$$Y_{ik} \sim \text{Cat}\left(\text{softmax}(s_{0ik}, \ldots, s_{jik})\right)$$
$$s_{0ik} = 0$$
$$s_{jik} = \beta_{0k} + \beta_{1k} Age_i + \beta_{2k} Female_i$$

$$\beta_{0k} = \gamma_{00} + \gamma_{01} TExp_k + \gamma_{02} TFemale_k + \eta_{0k}$$
$$\beta_{1k} = \gamma_{10} + \eta_{1k}$$
$$\beta_{2k} = \gamma_{20} + \gamma_{21} TFemale_k + \eta_{2k}$$

Ordinal `family = cumulative`

$$Y_{ik} \sim \text{Categorical}(\mathbf{p_{ik}})$$
$$p_{jik} = q_{jik} - q_{(j-1)ik}$$
$$\text{logit}(q_{jik}) = a_{jk} - \phi_{ik}$$
$$\phi_i = \beta_{1k} Age_i + \beta_{2k} Female_i$$

$$a_{jk} = \gamma_{00} + \gamma_{01} TExp_k + \gamma_{02} TFemale_k + \eta_{0k}$$
$$\beta_{1k} = \gamma_{10} + \eta_{1k}$$
$$\beta_{2k} = \gamma_{20} + \gamma_{21} TFemale_k + \eta_{2k}$$

Same model structure can be applied to any of the GLMs we've looked at.

Specifying these can be a pain, but normally you don't need to mess with the details — they're well implemented in `brms`. Simply specify the correct `family` parameter and multiple equations (multinomial) and intercepts (ordinal) will be created.

See the help for `brmsfamily` in R for the exhaustive list of model families available.