Mixed effects examples

Nick Graetz

October 2020

1 Four versions of the Week 4 unemployment-depression model

Model A) "unconditional growth model" (Week 4, slide 22)

$$Y_{ij} = \boxed{\beta_0} + \boxed{\beta_1}(T_{ij}) + \alpha_i + \gamma_i(T_{ij}) + \epsilon_{ij} \tag{1}$$

$$\begin{bmatrix} \alpha_i \\ \gamma_i \end{bmatrix} = N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \boxed{\sigma^2_\alpha} & \boxed{\sigma^2_{\alpha\gamma}} \\ \sigma^2_{\alpha\gamma} & \boxed{\sigma^2_\gamma} \end{bmatrix} \right) \tag{2}$$

$$\epsilon_{ij} = N(0, \boxed{\sigma^2_\epsilon}) \tag{3}$$

Yellow highlights your "fixed effects", or what we are treating as fixed unknown constants (β_0 is the global intercept, and β_1 is the global time effect or average time slope). The other parameters are your "random effects", or what we are treating as unknown random variables. Because these are random variables, they require specifying another equation to define their distributions. Here we are saying they are multivariate normal, so their joint distribution is defined by the variance parameters highlighted in green: the variance for the intercept random effect σ^2_α, the variance for the time random effect (σ^2_γ), and their covariance ($\sigma^2_{\alpha\gamma}$). In the R package `nlme`, running `summary()` on a fitted model object gives you these three terms as standard deviations, so you could square them to get the variances. Most software for mixed effect models, including `nlme`, automatically return summary statistics for the fixed effects (yellow terms above), fitted variance terms defining your random effects (green terms above), and the residual variance (orange term above). This is because you know the mean of all your random effects is 0. If you want the actual random effects themselves for every individual (α_i and γ_i above), you can do `random.effects()` on your fitted model object.

We can add unemployment status (U) to the fixed effects above to get Model B, and write Model B in the composite form above by simply adding one more yellow term (β_2):

$$Y_{ij} = \boxed{\beta_0} + \boxed{\beta_1}(T_{ij}) + \boxed{\beta_2}(U_{ij}) + \alpha_i + \gamma_i(T_{ij}) + \epsilon_{ij} \tag{4}$$

$$\begin{bmatrix} \alpha_i \\ \gamma_i \end{bmatrix} = N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \boxed{\sigma^2_\alpha} & \boxed{\sigma^2_{\alpha\gamma}} \\ \sigma^2_{\alpha\gamma} & \boxed{\sigma^2_\gamma} \end{bmatrix} \right) \tag{5}$$

$$\epsilon_{ij} = N(0, \boxed{\sigma^2_\epsilon}) \tag{6}$$

Or we can write this same Model B using a level-1/level-2 specification. Notice how we are just rearranging terms from the composite specification into multiple lines (Week 4, slide 23):

$$Y_{ij} = \beta_{0i} + \beta_{1i}(T_{ij}) + \beta_{2i}(U_{ij}) + \epsilon_{ij} \tag{7}$$

$$\beta_{0i} = \boxed{\beta_0} + \alpha_i \tag{8}$$

$$\beta_{1i} = \boxed{\beta_1} + \gamma_i \tag{9}$$

$$\beta_{2i} = \boxed{\beta_2} \tag{10}$$

$$\begin{bmatrix} \alpha_i \\ \gamma_i \end{bmatrix} = N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \boxed{\sigma^2_\alpha} & \boxed{\sigma^2_{\alpha\gamma}} \\ \sigma^2_{\alpha\gamma} & \boxed{\sigma^2_\gamma} \end{bmatrix} \right) \tag{11}$$

$$\epsilon_{ij} = N(0, \boxed{\sigma^2_\epsilon}) \tag{12}$$

And we can fit Model A and Model B using `nlme` in R:

```
modelA <- lme(cesd~months, data=unemployment, random=~months|id, method="ML")

modelB <- lme(cesd~months+unemp, data=unemployment, random=~months|id,
method="ML")
```

In Model C, we interact time and unemployment in our fixed effects:

$$Y_{ij} = \boxed{\beta_0} + \boxed{\beta_1}(T_{ij}) + \boxed{\beta_2}(U_{ij}) + \boxed{\beta_3}(T_{ij} * U_{ij}) + \alpha_i + \gamma_i(T_{ij}) + \epsilon_{ij} \tag{13}$$

$$\begin{bmatrix} \alpha_i \\ \gamma_i \end{bmatrix} = N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \boxed{\sigma^2_\alpha} & \boxed{\sigma^2_{\alpha\gamma}} \\ \sigma^2_{\alpha\gamma} & \boxed{\sigma^2_\gamma} \end{bmatrix} \right) \tag{14}$$

$$\epsilon_{ij} = N(0, \boxed{\sigma^2_\epsilon}) \tag{15}$$

Model C as level-1/level-2:

$$Y_{ij} = \beta_{0i} + \beta_{1i}(T_{ij}) + \beta_{2i}(U_{ij}) + \beta_{3i}(T_{ij} * U_{ij}) + \epsilon_{ij} \tag{16}$$

$$\beta_{0i} = \boxed{\beta_0} + \alpha_i \tag{17}$$

$$\beta_{1i} = \boxed{\beta_1} + \gamma_i \tag{18}$$

$$\beta_{2i} = \boxed{\beta_2} \tag{19}$$

$$\beta_{3i} = \boxed{\beta_3} \tag{20}$$

$$\begin{bmatrix} \alpha_i \\ \gamma_i \end{bmatrix} = N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \boxed{\sigma^2_\alpha} & \boxed{\sigma^2_{\alpha\gamma}} \\ \sigma^2_{\alpha\gamma} & \boxed{\sigma^2_\gamma} \end{bmatrix} \right) \tag{21}$$

$$\epsilon_{ij} = N(0, \boxed{\sigma^2_\epsilon}) \tag{22}$$

Fit Model C in R:

```
modelC <- lme(cesd~months*unemp, data=unemployment, random=~months|id,
method="ML")
```

Notice above, we assume the main effect and the interactive effect of unemployment are fixed and do not vary by individual. Finally in Model D, we introduce a random effect to both of these terms, but remove the fixed main effect of time:

$$
\begin{aligned}
Y_{ij} &= \beta_0 + \beta_1(U_{ij}) + \beta_2(T_{ij} * U_{ij}) + \alpha_i + \zeta_i(U_{ij}) + \delta_i(T_{ij} * U_{ij}) + \epsilon_{ij} && (23)\\
\begin{bmatrix} \alpha_i \\ \zeta_i \\ \delta_i \end{bmatrix} &= N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\alpha} & \sigma^2_{\alpha\zeta} & \sigma^2_{\alpha\delta} \\ \sigma^2_{\zeta\alpha} & \sigma^2_{\zeta} & \sigma^2_{\zeta\delta} \\ \sigma^2_{\delta\alpha} & \sigma^2_{\delta\zeta} & \sigma^2_{\delta} \end{bmatrix} \right) && (24)\\
\epsilon_{ij} &= N(0, \sigma^2_{\epsilon}) && (25)
\end{aligned}
$$

Model D as level-1/level-2 (again, notice how we are just rearranging terms from the composite model):

$$
\begin{aligned}
Y_{ij} &= \beta_{0i} + \beta_{1i}(U_{ij}) + \beta_{2i}(T_{ij} * U_{ij}) && (26)\\
\beta_{0i} &= \beta_0 + \alpha_i && (27)\\
\beta_{1i} &= \beta_1 + \zeta_i && (28)\\
\beta_{2i} &= \beta_2 + \delta_i && (29)\\
\begin{bmatrix} \alpha_i \\ \zeta_i \\ \delta_i \end{bmatrix} &= N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\alpha} & \sigma^2_{\alpha\zeta} & \sigma^2_{\alpha\delta} \\ \sigma^2_{\zeta\alpha} & \sigma^2_{\zeta} & \sigma^2_{\zeta\delta} \\ \sigma^2_{\delta\alpha} & \sigma^2_{\delta\zeta} & \sigma^2_{\delta} \end{bmatrix} \right) && (30)\\
\epsilon_{ij} &= N(0, \sigma^2_{\epsilon}) && (31)
\end{aligned}
$$

Fit Model D in R:

```
   modelD <- lme(cesd unemp+unemp:months,
random= unemp+unemp:months|id,
data=unemployment,
control = list(opt = "optim", tolerance=1e-6))
```

Models A-D here are fit in R in the script I uploaded for the TA session, where we also visualized the outputs of the fixed effects (example_code.R under Module 5). These correspond to the results table from Week 4, slide 24. Notice how in that table in the slides, we only report the variance components from the random effects (the diagonal of the variance-covariance matrix above) rather than also including the covariance terms. We could also report these, but it can quickly get cluttered if those are not interesting for your research question.