* Return PS2

JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

Lecture 13

More with Linear Mixed Models

Lecture 14/15 : missing data
Lecture 16 : office hour
Lab 7 open session PS3
Lab 8 open session final project

Subject specific or random effects models

mixed models

- Consider the data generating structure within the NEPAL1 and NEPAL2 simulated datasets:
  - Children are enrolled between 1 and 5 months of age
  - Children are followed over time and growth in weight is recorded every 4 months for a total of 5 assessments (enrollment + 4 follow-ups)

- For each child, we can think of the child's growth:

$$Y_{ij} = \beta_{0i} + \beta_{1i} age_{ij} + \beta_{2i}(age_{ij} - 6)^{+} + \boxed{e_{ij}}$$

child specific regression model

residuals

distance between the child's observed weight and the expected weight for the child

Subject specific or random effects models

- The β describe characteristics of the specific children and we assume that these characteristics can vary from child to child, specifically,

residuals
random effects

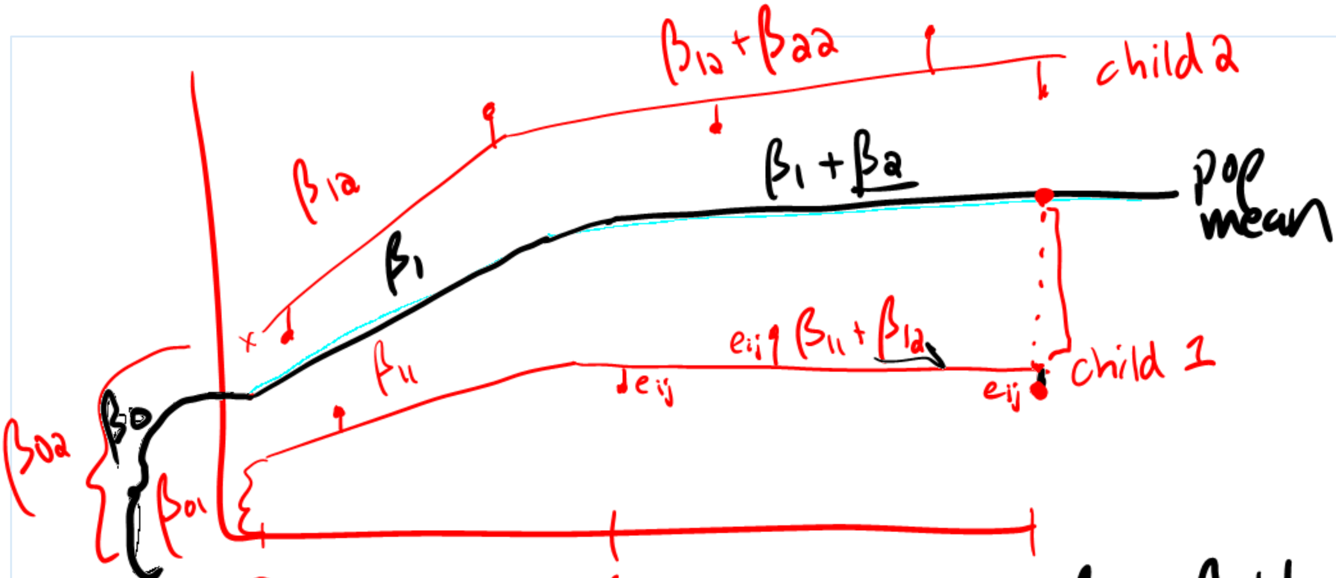
$$\underbrace{\begin{bmatrix}\beta_{0i}\\ \beta_{1i}\\ \beta_{2i}\end{bmatrix}}_{\text{children's growth}} = \underbrace{\begin{bmatrix}\beta_{0}\\ \beta_{1}\\ \beta_{2}\end{bmatrix}}_{\text{pop mean growth}} + \underbrace{\begin{bmatrix}b_{0i}\\ b_{1i}\\ b_{2i}\end{bmatrix}}$$

Cor(b_{0i}, b_{1i})

$$\beta_i = \beta + b_i, b_i \sim MVN(0, D), D = \begin{bmatrix}\tau_0^2 & \tau_{01} & \tau_{02}\\ \tau_{01} & \tau_1^2 & \tau_{12}\\ \tau_{02} & \tau_{12} & \tau_2^2\end{bmatrix}$$

variation across children in characteristics of growth

Visualization



for child 1: $\beta_{01} = \beta_0 + \underbrace{b_{01}}_{\text{negative}}$

$\beta_{11} = \beta_1 + \underbrace{b_{11}}_{\text{negative}}$

for child 2: $\beta_{02} = \beta_0 + \underbrace{b_{02}}_{\text{positive}}$

$\beta_{12} = \beta_1 + \underbrace{b_{12}}_{\text{positive}}$

General Model

We can rewrite the model above as:

β_{0i} β_{1i} β_{2i}

$$Y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})age_{ij} + (\beta_2 + b_{2i})(age_{ij} - 6)^+ + e_{ij}$$

In vector notation,

$$Y_{ij} = \underbrace{\begin{bmatrix} 1 \\ age_{ij} \\ (age_{ij} - 6)^+ \end{bmatrix}' \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}}_{\text{pop mean}} + \underbrace{\begin{bmatrix} 1 \\ age_{ij} \\ (age_{ij} - 6)^+ \end{bmatrix}' \begin{bmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \end{bmatrix} + e_{ij}}_{\text{error variation}}$$

Even more generally,

$$Y_{ij} = X_{ij}'\beta + Z_{ij}'b_i + e_{ij}$$

where $b_i \sim MVN(0, D)$, e_{ij} iid $N(0, \sigma^2)$ and b_i and e_{ij} are independent!

at most Z contains

1, age_{ij}, $(age_{ij} - 6)^+$

Means and Variances

- In the random effects model, we express the mean function for an individual subject as:

$$E(Y_{ij}|X_{ij}, b_i) = X_{ij}\beta + Z_{ij}b_i$$

→ child specific trajectories

- We can express the population mean (i.e. the average over all subjects) as:

$$E(Y_{ij}|X_{ij}) = E[E(Y_{ij}|X_{ij}, b_i)] = E[X_{ij}\beta + Z_{ij}b_i] = X_{ij}\beta$$

pop means

- We can derive the variance of Y_{ij} as

$$Var(Y_{ij}|X_{ij}) = E_{b_i}[Var(Y_{ij}|X_{ij}, b_i)] + Var_{b_i}[E(Y_{ij}|X_{ij}, b_i)]$$

$$Var(Y_{ij}|X_{ij}) = E_{b_i}[\sigma^2] + Var_{b_i}[X'_{ij}\beta + Z'_{ij}b_i]$$

$$Var(Y_{ij}|X_{ij}) = \sigma^2 + \underbrace{Z'_{ij}DZ_{ij}}$$

allow the variance to depend on elements of $Z_{ij} \Rightarrow X_{ij}$

Correlation

- Assume a random intercept only model:
  - $Y_{ij} = \beta_{0i} + \beta_1 age_{ij} + \beta_2 (age_{ij} - 6)^+ + e_{ij}, \beta_{0i} \sim N(\beta_0, \tau_0^2), e_{ij} \sim N(0, \sigma^2), Cov(\beta_{0i}, e_{ij}) = 0$
  - $Y_{ij} = \beta_0 + b_{0i} + \beta_1 age_{ij} + \beta_2 (age_{ij} - 6)^+ + e_{ij}, b_{0i} \sim N(0, \tau_0^2), e_{ij} \sim N(0, \sigma^2), Cov(b_{0i}, e_{ij}) = 0$
  - $Corr(Y_{ij}, Y_{ik}) = \frac{\tau_0^2}{\tau_0^2 + \sigma^2}$ exchangeable model

- $\boxed{\text{Assume a random intercept and random slope for age model:}}$
  - $Y_{ij} = \beta_0 + b_{0i} + (\beta_1 + b_{i1}) age_{ij} + \beta_2 (age_{ij} - 6)^+ + e_{ij}$, *where*

$b_{0i} \sim N(0, \tau_0^2), b_{1i} \sim N(0, \tau_1^2), Cov(b_{0i}, b_{1i}) = \tau_{01}, e_{ij} \sim N(0, \sigma^2), Cov(b_{0i}, e_{ij}) = 0, Cov(b_{1i}, e_{ij}) = 0$

  - $Corr(Y_{ij}, Y_{ik}) =$ $f(\tau_0^2, \tau_1^2, \tau_{01}, \sigma^2, age_{ij}, age_{ik})$

Revisit NEPAL1 analysis AND do another example, NEPAL2