JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

* PS 2 solutn / grades
→ email @ PS3 / quiz deadline
→ Final project Friday

Lecture 13

More with Linear Mixed Models

Lectures 14/15 missing data

Lab 7/8 open sessions

[Lecture 16]

Subject specific or random effects 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:

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

Child specific

2 Frameworks

↳ marginal models

↳ subject specific
conditional
model

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

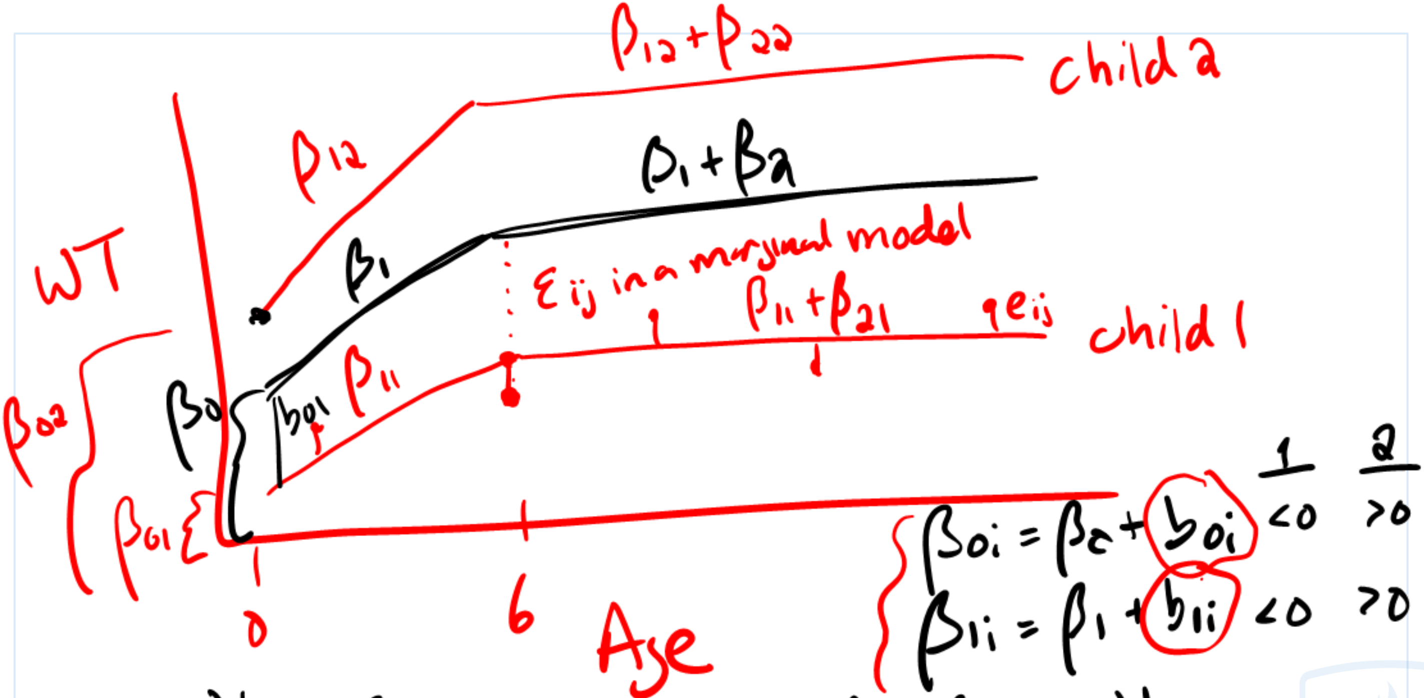
random effects

$$\begin{bmatrix} \beta_{0i} \\ \beta_{1i} \\ \beta_{2i} \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \end{bmatrix}$$

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

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

Visualization



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

General Model

We can rewrite the model above as:

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

In vector notation,

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

Even more generally,

pop mean

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

+ variance

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

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's specific prediction

- 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 mean

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

$$Var(Y_{ij}|X_{ij}) = E_{b_i}[Var(Y_{ij}|X_{ij}, b_i)] + Var_{b_i}[E(Y_{ij}|X_{ij}, b_i)]$$
$$Var(Y_{ij}|X_{ij}) = E_{b_i}[\sigma^2] + Var_{b_i}[X_{ij}'\beta + Z_{ij}'b_i]$$
$$Var(Y_{ij}|\underline{X_{ij}}) = \sigma^2 + Z_{ij}'DZ_{ij}$$

within subject variance

between subject

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

- 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*

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

  - $Corr(Y_{ij}, Y_{ik}) =$ $f(age_{ij}, age_{ik})$

Revisit NEPAL1 analysis AND do another example, NEPAL2