Correlated / Clustered / Longitudinal Data — 4/14/22

n "clusters" $i = 1, 2, \ldots, n$ t_i = index of t^{th} measurement on cluster i

$$\underline{Y_i} = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{it_i} \end{bmatrix} \qquad Cov(\underline{Y_i}) = \begin{bmatrix} Var(Y_{i1}) & Cov(Y_{i1}, Y_{i2}) & \cdots \\ Cov(Y_{i2}, Y_{i1}) & Var(Y_{i2}) & \\ \vdots & & \ddots \\ Cov(Y_{it_i}, Y_{i1}) & & Var(Y_{it_i}) \end{bmatrix}$$

GEE → Model 1^{st} two moments —

$E(\underline{Y_i})$ and $Cov(\underline{Y_i})$ — no distributional assumption

primary interest

nuisance parameter

→ Robust to choice of covariance structure

Common Correlation Structures: $Corr(\underline{Y_i}) = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \cdots \\ \rho_{21} & 1 & & \\ \vdots & & \ddots & \\ \rho_{2t_i} & & & 1 \end{bmatrix}$

Independence: $Corr(\underline{Y_i}) = I_{t_i}$

$$= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \\ \vdots & & \ddots & \\ & & & 1 \end{bmatrix}$$

$t_i = 3$

Exchangeable / Compound Symmetry: $Corr(\underline{Y_i}) = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}$

AR(1) — autoregressive:

$$Corr(\underline{Y_i}) = \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix}$$

Unstructured: $Corr(\underline{Y_i}) = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}$

Option 2: Subject-specific models – add random / fixed effect per cluster

Partial table for each subject:

Example

		Asthma Yes	Asthma No
Age	13	1	0
	20	0	1

2 ← total # observations

= # repeated measures

	1	0
	1	0

	0	1
	1	0

	0	1
	0	1

Model: $$\text{logit}(P(Y_{it}=1)) = \underbrace{\alpha_i}_{\mu + \tau_i} + \beta x_{it}$$

$\alpha_i \sim N(\mu, \sigma^2)$

$\tau_i \sim N(0, \sigma^2)$

person i at time t

Compare to GEE: $$\text{logit}(P(Y_t = 1)) = \alpha + \delta x_t \qquad t = 1, 2$$

Correlation within subject is induced

Example - Asthma

GEE: $Y_t = \begin{cases} 1 & \text{asthma at time } t \\ 0 & \text{else} \end{cases}$

Assumptions:

$E(Y_i)$

$$\text{logit}(P(Y_t = 1)) = \alpha + \delta x_t$$

$$x_t = \begin{cases} 1 & \text{age } 20 \\ 0 & \text{age } 13 \end{cases}$$

$$\text{Cov}\left(\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}\right) = \begin{bmatrix} \sigma^2 & \tau \\ \tau & \sigma^2 \end{bmatrix}$$

GLMM: $Y_{it} = \begin{cases} 1 & \text{asthma at time } t \text{ for person } i \\ 0 & \text{else} \end{cases}$

$$\text{logit}(P(Y_{it} = 1)) = \mu + \alpha_i + \beta x_{it}$$

$$\alpha_i \overset{iid}{\sim} N(0, \sigma^2)$$

$$Y_{it} \mid \alpha_i \sim \text{Bin}(1, \underbrace{E(Y_{it} \mid \alpha_i)}_{\pi_{it}})$$

$$\text{logit}(E(Y_{it} \mid \alpha_i)) = \mu + \alpha_i + \beta x_{it}$$

$0 \le \quad \le 1$

Aside: Identity link:

$$E(Y_{it} \mid \alpha_i) = \mu + \alpha_i + \beta x_{it}$$

$$\Rightarrow E(Y_{it}) = E[E(Y_{it} \mid \alpha_i)] = \mu + E(\alpha_i) + \beta x_{it} = \mu + \beta x_{it}$$

Issue: $g(E(Y_{it} \mid \alpha_i)) = \mu + \alpha_i + \beta x_{it}$ ← GLMM

$$g(E(Y_{it})) \neq E[g(E(Y_{it} \mid \alpha_i))]$$

$$g(E(E(Y_{it} \mid \alpha_i)))$$

Fitted Models:

GEE:

$$\text{logit}(\hat{P}(Y_t=1)) = -1.782 - 0.248 X_t$$

GLMM:

$$\text{logit}(\hat{P}(Y_{it}=1 \mid \alpha_i)) = -8.087 + \hat{\alpha}_i - 1.032 X_{it}$$

Odds Ratio $\left\{ P(\text{asthma} \mid \text{age}_{20}) \,/\, P(\text{no asthma} \mid \text{age}_{20}) \right\} \Big/ \dfrac{P(\text{asthma} \mid \text{age}_{13})}{P(\text{no asthma} \mid \text{age}_{13})}$

GEE: $e^{-0.248} = 0.780$ $\Downarrow$ Estimated 22% decrease in the odds of asthma at age 20 compared to age 13.

population-averaged

GLMM: $e^{-1.032} = 0.356$ $\Downarrow$ Estimated 64% decrease ...

subject-specific

Simple case $\longrightarrow$ closed form solutions to those estimated odds ratios

Original data:

		Age 20	
		Asthma	Not
Age 13	Asthma	$n_{11}=50$	$n_{12}=22$
	No	$n_{21}=8$	n_{22} 420

GEE:

$$\hat{OR}_{\text{age}13/20} = \frac{ad}{bc} = 1.282$$

$$\Rightarrow \text{age}20/\text{age}13 = \frac{1}{1.282} = 0.780$$

Create marginal table:

		Asthma Yes	No
Age	13	$a=72$	$b=428$
	20	$c=58$	$d=442$

GLMM: $\hat{OR} = n_{12}/n_{21}$

$$= 22/8 = 2.75$$

$$\longrightarrow \frac{\text{age }20}{\text{age }13} = \frac{1}{2.75} = 0.364$$