Generalized Linear Models for Longitudinal Data

- Generalized Linear Models: A Review
- The Logistic regression model

  Marginal model

  Random effects model

  Transition Model
- Contrasting Approaches

GLM examples

- **Linear regression**

$$\mu_i = \boldsymbol{X}_i\boldsymbol{\beta};\ g(\mu_i) = \mu_i$$
$$Y_i \sim N(\mu_i, \sigma^2)$$

- **Logistic regression**

$$\log\left(\frac{\mu_i}{1-\mu_i}\right) = \boldsymbol{X}_i\boldsymbol{\beta};\ g(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right)$$
$$Y_i \sim \text{Bernoulli}(\mu_i)$$

- **Poisson regression**

$$\log\mu_i = \boldsymbol{X}_i\boldsymbol{\beta};\ g(\mu_i) = \log\mu_i$$
$$Y_i \sim \text{Poisson}(\mu_i)$$

Note for GLMs

- varY_i may be a function of μ_i

\- Logistic: var$Y_i = \mu_i(1-\mu_i)$

\- Poisson: var$Y_i = \mu_i$

Logistic regression model

- Marginal Logistic Regression Model
- Logistic Model with Random Effects
- Transition Logistic Model

Marginal Logistic regression model

- Goal: to assess the dependence of respiratory infection on vitamin A status in the Indonesian Children's Health Study
- $x_{ij} = 1$ if child i is vitamin A deficient at visit j
- $y_{ij} = 1$ child i has respiratory infection at visit j
- $\mu_{ij} = E(Y_{ij}) = P(Y_{ij} = 1)$
- logit$\mu_{ij} = \beta_0 + \beta_1 x_{ij}$
- $P(Y_{ij} = 1) = \frac{\exp(\beta_0+\beta_1 x_{ij})}{1+\exp(\beta_0+\beta_1 x_{ij})}$
- $var(Y_{ij}) = \mu_{ij}(1-\mu_{ij})$
- $corr(Y_{ij}, Y_{ik}) = \alpha$

Parameter Interpretation

- $\frac{\exp(\beta_0)}{1+\exp(\beta_0)} = P(Y_{ij} = 1 \mid x_{ij} = 0)$ *frequency of infected children among the subpopulation that* **is not** *vitamin A deficiency*
- $\frac{\exp(\beta_0+\beta_1)}{1+\exp(\beta_0+\beta_1)} = P(Y_{ij} = 1 \mid x_{ij} = 1) =$ *frequency of infected children among the subpopulation that* **is** *vitamin A deficiency*
- $e^{\beta_0} = \frac{P(Y_{ij}=1|x_{ij}=0)}{Pr(Y_{ij}=0|x_{ij}=0)}$ *ratio (odds) of the frequency of infected to uninfected children among the subpopulation that* **is not** *vitamin A deficiency*
- $e^{\beta_0+\beta_1} = \frac{P(Y_{ij}=1|x_{ij}=1)}{Pr(Y_{ij}=0|x_{ij}=1)}$ *ratio (odds) of the frequency of infected to uninfected children among the subpopulation that* **is** *vitamin A deficiency*
- $e^{\beta_1} = \frac{\exp(\beta_0+\beta_1)}{\exp(\beta_0)} =$ *odds of infection among vitamin A deficient children divided by the odds among children replete with vitamin A*

Correlation between binary outcomes

1. corr$(Y_{ij}, Y_{ik}) = \alpha$
2. better to model association among binary data using the odds ratio

$$OR(Y_{ij}, Y_{ik}) = \frac{P(Y_{ij}=1, Y_{ik}=1)P(Y_{ij}=0, Y_{ik}=0)}{P(Y_{ij}=1, Y_{ik}=0)P(Y_{ij}=0, Y_{ik}=1)}$$

Logistic Model with Random Effects

- We assume that the propensity for respiratory infection varies across children, reflecting their different genetic predispositions and unmeasured influences of environmental factors
- We assume that each child has his/her own propensity for respiratory disease $\beta_0^* + U_i$ but that the effect of vitamin A deficient β_1^* on this probability is the same for every child

$$\begin{aligned} \text{logit} P(Y_{ij} = 1 \mid U_i) &= (\beta_0^* + U_i) + \beta_1^* x_{ij} \\ U_i &\sim N(0, v^2) \end{aligned}$$

- Given i, we further assume that the repeated observations for the $i-th$ child are independent of one another
- $\beta_0^* =$ *log-odds of respiratory infection for a typical child with random effect $U_i = 0$*

- $\exp(\beta_1^*) =$ *odds of infection* **for a child – with random effect** U_i *when he/she is vitamin A deficient relative to when the same child is not*

$$\begin{aligned} \exp(\beta_1^*) &= \frac{\exp(\beta_0^* + U_i + \beta_1^*)}{\exp(\beta_0^* + U_i)} \\ &= \frac{P(Y_{ij}=1|U_i, x_{ij}=1)/P(Y_{ij}=0|U_i, x_{ij}=1)}{P(Y_{ij}=1|U_i, x_{ij}=0)/P(Y_{ij}=0|U_i, x_{ij}=0)} \end{aligned}$$

- $v^2 =$ *degree of heterogeneity across children in the propensity of disease, not attributable to x*

RE models: basic ideas

- There is a natural heterogeneity across individuals in their regression coefficients and that this heterogeneity can be explained by a probability distribution
- RE model most useful when the objective is to make inference about individuals rather than population average
- β_1^* represents the effects of the explanatory variables on an individual child's chance of infection
- *this is in contrast with the marginal model coefficients which describe the effect of explanatory variable on the population average*

Parameter Interpretation of a Logistic Regression Model with Random Effects

- logit$P(Y_{ij} = 1 \mid U_i, x_{ij} = 1) = \beta_0^\star + U_i + \beta_1^\star$
- logit$P(Y_{ij} = 1 \mid U_i, x_{ij} = 0) = \beta_0^\star + U_i$
- $Od(U_i, x_{ij} = 1) = \frac{P(Y_{ij}=1|U_i,x_{ij}=1)}{P(Y_{ij}=0|U_i,x_{ij}=1)} = \exp(\beta_0^\star + U_i + \beta_1^\star)$
- $Od(Ui, x_{ij} = 0) = \frac{P(Y_{ij}=1|U_i,x_{ij}=0)}{P(Y_{ij}=0|U_i,x_{ij}=0)} = \exp(\beta_0^\star + U_i)$
- $Od(U_i, x_{ij} = 1) = e^{\beta_1^\star} \times Od(U_i, x_{ij} = 0)$
- the odds of respiratory infection for a hypothetical child with random effect U_i and with vitamin A deficiency are equal to $e^{\beta_1^\star}$ times the odds of respiratory infection for the same hypothetical child with random effect U_i without vitamin A deficiency.

Transition Model for Binary Responses

- logit $P(Y_{ij} \mid y_{ij-1}) = \beta_0^{**} + \beta_1^{**} x_{ij} + \alpha y_{ij-1}$

  the chance of respiratory infection at time t_{ij} depends on explanatory variables but also on whether or not the child had infection 3 months earlier

  β_1^{**} = *change* $\times$ *unit change in* x *in the log odds of infection, among children with outcome* y_{ij-1} *at the prior visit*

- $e^\alpha = \frac{\exp(\beta_0^{**}+\beta_1^{**}x_{ij}+\alpha)}{\exp(\beta_0^{**}+\beta_1^{**}x_{ij})}$ = ratio of the odds of infection among children who did and did not have infection at the prior visit

$$\frac{P(Y_{ij}=1|Y_{ij-1}=1)}{P(Y_{ij}=0|Y_{ij-1}=1)} = \exp(\beta_0^{**} + \beta_1^{**} x_{ij} + \alpha)$$
$$\frac{P(Y_{ij}=1|Y_{ij-1}=0)}{P(Y_{ij}=0|Y_{ij-1}=0)} = \exp(\beta_0^{**} + \beta_1^{**} x_{ij})$$

Transition matrix

The logistic regression above can be specified as a *transition matrix*

$$\begin{array}{cc|cc} & & y_{ij} & \\ & & 0 & 1 \\ y_{ij-1} & 0 & \frac{1}{1+\exp(\beta_0^{**}+\beta_1^{**}x_{ij})} & \frac{\exp(\beta_0^{**}+\beta_1^{**}x_{ij})}{1+\exp(\beta_0^{**}+\beta_1^{**}x_{ij})} \\ & 1 & \frac{1}{1+\exp(\beta_0^{**}+\beta_1^{**}x_{ij}+\alpha)} & \frac{\exp(\beta_0^{**}+\beta_1^{**}x_{ij}+\alpha)}{1+\exp(\beta_0^{**}+\beta_1^{**}x_{ij}+\alpha)} \end{array}$$

Parameter Interpretation of a Transition Logistic Regression Model

- logit$P(Y_{ij} = 1 \mid y_{ij-1}, x_{ij} = 1) = \beta_0^{\star\star} + \beta_1^{\star\star} + \alpha y_{ij-1}$
- logit$P(Y_{ij} = 1 \mid y_{ij-1}, x_{ij} = 0) = \beta_0^{\star\star} + \alpha y_{ij-1}$
- $Od(y_{ij-1}, x_{ij} = 1) = \frac{P(Y_{ij}=1|y_{ij-1},x_{ij}=1)}{P(Y_{ij}=0|y_{ij-1},x_{ij}=1)} = \exp(\beta_0^{\star\star} + \alpha y_{ij-1} + \beta_1^{\star\star})$
- $Od(y_{ij-1}, x_{ij} = 0) = \frac{P(Y_{ij}=1|y_{ij-1},x_{ij}=0)}{P(Y_{ij}=0|y_{ij-1},x_{ij}=0)} = \exp(\beta_0^{\star\star} + \alpha y_{ij-1})$
- $Od(y_{ij-1}, x_{ij} = 1) = e^{\beta_1^{\star\star}} \times Od(y_{ij-1}, x_{ij} = 0)$
- the odds of respiratory infection for a hypothetical child with outcome at the previous visit equal to y_{ij-1} with vitamin A deficiency are equal to $e^{\beta_1^{\star\star}}$ times the odds of respiratory infection for the same hypothetical child with outcome at the previous visit equal to y_{ij-1} without vitamin A deficiency.

In summary

1. Marginal model

$$\text{logit}P(Y_{ij} = 1) = \beta_0 + \beta_1 x_{ij}$$

   β_1 *describes the effect of explanatory variables on the chance of infection in the* **entire population**

2. Random effects model

$$\text{logit}P(Y_{ij} = 1 \mid U_i) = \beta_0^* + U_i + \beta_1^* x_{ij}$$

   β_1^* *describes the effect of the explanatory variables on an* **individual** *chance of infection*

3. Transition model

$$\text{logit}P(Y_{ij} = 1 \mid y_{ij-1}) = \beta_0^{**} + \beta_1^{**} x_{ij} + \alpha y_{ij-1}$$

   β_1^{**} *describes the effect of explanatory variables on the chance of infection adjusted by the outcome for respiratory infection at the* **prior visit** Note that:

$$\text{logit}P(Y_{ij} = 1 \mid y_{ij-1} = 0) = \beta_0^{**} + \beta_1^{**} x_{ij}$$

   this is a Marginal logistic regression model for the sub-population of children that are free of infection at the previous visit

Contrasting Approaches

- In linear models, the interpretation of $\boldsymbol{\beta}$ is essentially independent of the correlation structure
- In non-linear models for discrete data, such us logistic regression, different assumptions about the source of correlation can lead to regression coefficients with distinct interpretations
- Two examples:
  Infant growth
  Respiratory Disease data

Linear Regression Model for infant growth

- Y_{ij} is the weight at age t_{ij}
- i is the child, j is the visit

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + \epsilon_{ij}$$

1. $E(Y_{ij}) = \beta_0 + \beta_1 t_{ij}$

   i.e.- the average weight for all infants in the population at any time t is $\beta_0 + \beta_1 t$

- β_1 *is the change per month in the population-average weight*

2. $corr(\epsilon_{ij}, \epsilon_{ik}) = \rho(t_{ij}, t_{ik}, \boldsymbol{\alpha})$

   $\rho(t_{ij}, t_{ik}, \boldsymbol{\alpha}) = \alpha_0$ if $\mid t_{ij} - t_{ik} \mid < 6$

   $\rho(t_{ij}, t_{ik}, \boldsymbol{\alpha}) = \alpha_1$ if $\mid t_{ij} - t_{ik} \mid \geq 6$

Linear Random Effects Model

$$\begin{array}{lcl} Y_{ij} & = & \beta_0^* + U_{i0} + (\beta_1^* + U_{i1})t_{ij} + Z_{ij} \\ Z_{ij} & \sim & N(0, \sigma^2) \\ (U_{i0}, U_{i1}) & \sim & N\left(0, \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}\right) \end{array}$$

In the linear random effects model, the regression coefficients also have a marginal interpretation

$$E[Y_{ij}] = \beta_0 + \beta_1 t_{ij}$$

- this is because the *average* of the rates of growth for individuals is the same as the change in the *population average* weight across time in a linear model
- if $G_{22} = 0 \Rightarrow corr(Y_{ij}, Y_{ik}) = \frac{G_{11}}{\sigma^2 + G_{11}}$

A Transition Model for infant growth

$$\begin{array}{lcl} Y_{ij} & = & \beta_0^{**} + \beta_1^{**} t_{ij} + \epsilon_{ij} \quad (1) \\ \epsilon_{ij} & = & \alpha \epsilon_{ij-1} + Z_{ij},\ Z_{ij} \sim N(0, \sigma^2) \quad (2) \end{array}$$

$$Y_{ij} = \beta_0^{**} + \beta_1^{**} t_{ij} + \alpha(y_{ij-1} - \beta_0^{**} - \beta_1^{**} t_{ij-1}) + Z_{ij}$$

$$E[Y_{ij} \mid y_{ij-1}] = \beta_0^{**} + \beta_1^{**} t_{ij} + \alpha(y_{ij-1} - \beta_0^{**} - \beta_1^{**} t_{ij-1})$$

$$(1)(2) \Rightarrow E[Y_{ij}] = \beta_0^{**} + \beta_1^{**} t_{ij}$$

this form of transition model, has coefficients which have a marginal interpretation

In summary

1. Marginal model $E[Y_{ij}] = \beta_0 + \beta_1 x_{ij}$

   β_1 describes the change in the average response for a unit change in x_{ij} for the **entire population**

2. Random effects model

   $$\begin{array}{lcl} E(Y_{ij} \mid U_i) & = & \beta_0^* + U_i + \beta_1^* x_{ij} \\ E[Y_{ij}] & = & E[E(Y_{ij} \mid U_i)] = \beta_0^* + E[U_i] + \beta_1^* x_{ij} = \beta_0^* + \beta_1^* x \end{array}$$

   β_1^* describes the change the the average response for a unit change in x_{ij} for a **particular subject**

3. Transition Regression model

   $$\begin{array}{lcl} E(Y_{ij} = 1 \mid y_{ij-1}) & = & \beta_0^{**} + \beta_1^{**} x_{ij} + \alpha y_{ij-1} \\ E[Y_{ij}] & = & E[E(Y_{ij} = 1 \mid y_{ij-1})] = \beta_0^{**} + \beta_1^{**} x_{ij} \end{array}$$

   β_1^{**} describes the change of the average response for a unit change in x_{ij} **among subjects with y_{ij-1} in the prior visit**

$$\boxed{\boldsymbol{\beta} = \boldsymbol{\beta}^* = \boldsymbol{\beta}^{**}}$$

A Simulation Example

1. Set the parameters $\beta_0^* = -2,\ \beta_1^* = 0.4,\ v^2 = 2$
2. generate $u_i \sim N(0, v^2)$
3. for u_i calculate $P(Y_{ij} = 1 \mid u_i, x_{ij} = 0)$ and $P(Y_{ij} = 1 \mid u_i, x_{ij} = 1)$, i.e. the chance of infection when the child is vitamin A deficient and when he is not

- for example, a child with $u_i = 0$ has 17% chance of infection when is vitamin A deficient, and 12% chance of infection when not:

  $P(Y_{ij} = 1 \mid u_i = 0, x_{ij} = 0) = \frac{\exp(\beta_0^*)}{1+\exp(\beta_0^*)} = 0.12$

  $P(Y_{ij} = 1 \mid u_i = 0, x_{ij} = 1) = \frac{\exp(\beta_0^*)}{1+\exp(\beta_0^*)} \times \exp(\beta_1^*) = 0.17$
- odds ratio for vitamin A deficiency is the some for every child, equal to $\exp(\beta_1^*) = 1.5$
- for example, a child with $u_i = 2$ has 74% chance of infection when is vitamin A deficient and 50% chance of infection when not

$P(Y_{ij}=1 \mid u_i=2, x_{ij}=0) = \frac{\exp(\beta_0^*+2)}{1+\exp(\beta_0^*+2)} = 0.5$

$P(Y_{ij}=1 \mid u_i=2, x_{ij}=0) = \frac{\exp(\beta_0^*)+2}{1+\exp(\beta_0^*+2)} \times \exp(\beta_1^*) = 0.74$

- However the corresponding change in absolute risk differs depending on the baseline rate
- the population rate of infection is the average risk, which is given by $P(Y_{ij}=1) = \int P(Y_{ij}=1 \mid U_i) f(u_i, v^2) dU_i$

Implementation of Simulation study

1. generate $U_i \sim N(0, v^2)$ where $v=2$
2. generate $Y_{ij} \sim$ Bernouilli(μ_i) where
3. $\mu_i = \frac{\exp(\beta_0^*+U_i+\beta_1^* x_{ij}}{1+exp(\beta_0^*+U_i+\beta_1^* x_{ij})}$
4. estimate the fraction of people in the population with (and without vitamin A deficiency)
5. $\hat{P}(Y_{ij}=1 \mid x_{ij}=1) = 0.18$
6. $\hat{P}(Y_{ij}=1 \mid x_{ij}=0) = 0.23$
7. find β_0 and β_1 and compare them with β_0^* and β_1^*

If we use a marginal model

- in the marginal model, we ignore the differences among children and model the population-average

$$P(Y_{ij}=1) \qquad \text{rather than} \qquad P(Y_{ij}=1 \mid U_i)$$

- We estimate the fraction of people with infection among the people with(without) vitamin A deficiecy

$\hat{P}(Y_{ij}=1 \mid x_{ij}=0) = 0.18$ for vitamin A replete

$\hat{P}(Y_{ij}=1 \mid x_{ij}=1) = 0.23$ for vitamin A deficient

- infection rate in the sub-group that has sufficient vitamin A deficiency is $\exp(\beta_0)/(1+\exp(\beta_0)) = 0.18$ so that $\beta_0 = -1.51$
- Odds ratio for vitamin A deficiency is

$$\exp(\beta_1) = \frac{P(Y_{ij}=1|x_{ij}=1)/P(Y_{ij}=0|x_{ij}=1)}{P(Y_{ij}=1|x_{ij}=0)/P(Y_{ij}=0|x_{ij}=0)}$$

$$= \frac{(0.23)/(1-0.23)}{0.18/(1-0.18)} = 1.36$$

so that $\beta_1 = 0.31$

Marginal Model vs Random effects

- *Marginal and Random Effects model parameters differ in the logistic model*
- **Marginal:** ratio of population odds
- **Random Effects:** ratio of individual's odds
- Marginal parameter values are smaller in absolute values than their random effects analogue
- Parameters in transition models also differ from either the random effects and marginal model parameters.

$$\boxed{\mid \beta_k \mid \leq \mid \beta_k^* \mid}$$