

Summary of Modeling Strategies

Modeling of the Mean Responses

- Marginal models:

- The marginal expectation of the response, $E(Y_{ij}) = \mu_{ij}$, depends on the covariates, \mathbf{x}_{ij} , through a link function g ,

$$E(Y_{ij}) = \mu_{ij} = g^{-1}(\mathbf{x}_{ij}^T \boldsymbol{\beta}).$$

- Random effects (conditional) models:

- Conditional on subject specific, unobserved random variables \mathbf{b}_i ,

$$E(Y_{ij} | \mathbf{b}_i) = \mu_{ij}^* = g^{-1}(\mathbf{x}_{ij}^T \boldsymbol{\beta}^* + \mathbf{z}_{ij}^T \mathbf{b}_i).$$

- Transition models:

- Let $\mathcal{H}_{ij} = (Y_{i1}, \dots, Y_{ij-1})$ denote the history of Y_{ij} ,

$$E(Y_{ij} | \mathcal{H}_{ij}) = \mu_{ij}^{**} = g^{-1} \left\{ \mathbf{x}_{ij}^T \boldsymbol{\beta}^{**} + \sum_{r=1}^s f_r(\mathcal{H}_{ij}, \boldsymbol{\alpha}) \right\}.$$

Modeling of the Covariances

- Marginal models:

- The marginal variance of the response depends on the marginal mean,

$$\text{Var}(Y_{ij}) = \phi v(\mu_{ij}),$$

where v is a known function and the scale parameter ϕ may also depend on some covariates.

- The correlation between Y_{ij} and Y_{ik} is a function of the marginal mean:

$$\text{Cor}(Y_{ij}, Y_{ik}) = \rho(\mu_{ij}, \mu_{ik}, \boldsymbol{\alpha}),$$

where ρ is a known function and the correlation parameters $\boldsymbol{\alpha}$ may depend on covariates.

- Random effects (conditional) models:
 - Typically, conditioning on the random effects, \mathbf{b}_i , the responses Y_{i1}, \dots, Y_{in_i} are *independent* with *exponential family* distribution.
 - The random effects \mathbf{b}_i have mean $\mathbf{0}$ and variance D .
 - Typically the random effects are assumed to be multivariate Gaussian.
- Transition models:
 - Typically, $Y_{ij} | \mathcal{H}_{ij}$ is assumed to have an *exponential family* distribution with variances

$$\text{Var}(Y_{ij} | \mathcal{H}_{ij}) = \phi v(\mu_{ij}^{**}).$$

Generalized estimating equations (GEEs) vs maximum likelihood (ML)

- Advantages of GEE:
 - No need to specify a complete probability model (the likelihood).
 - Robust to misspecification of the correlation model (the marginal parameter estimates are still consistent).
- Disadvantages of GEE:
 - Can't use likelihood-based model selection criteria (LRT, AIC, BIC).
 - Less efficient when a relatively good probability model is available.
- GEE is typically used for marginal models whereas ML is typically used for conditional and transition models. In general, either method can be used in any of the modeling approaches.

Choices of Models

- The conditional and transition models can, in principle, be “marginalized” by integrating out the random effects or the history.
- In general, the regression parameters β , β^* and β^{**} have different interpretations and not directly comparable.
- The choice of model should depend on the scientific question of interest.

Interpretation

- Marginal model: β is the mean (expected) difference in the response for the two populations (individuals) with the identical covariate values but differ in X by one unit.
- Conditional model: β^* is the expected difference in the response for two individuals with the identical covariate values and *identical random effects* but differ in X by one unit.
 - It is somewhat artificial to have two individuals with the same random effects but interpretation of between-subject covariate effects (gender, treatment, etc.) is difficult otherwise.
 - For within-subject covariate effects (age, time), it is perhaps more natural to interpret it as the expected difference for the *same* individual when its X changes by one unit while other covariates (including random effects) being held as the same.
 - Alternatively, we can interpret β^* conditional on the random effects being 0, that is, the effect of the variable X on an *average* individual.
 - Gaussian case: $\beta = \beta^*$
 - Poisson case (log link): $\beta = \beta^*$ for non-intercept predictor and when only the random intercept is present.
 - Binary case (logistic link): $\beta \neq \beta^*$. β^* is usually larger than β because it less noisy when using its own observation as control.
- Transitional models: β^{**} is the effect of X while other covariates and the history are held equal.
 - β^{**} is potentially misleading to study relationship between Y and X .

Binary Response

Marginal Model

Consider a single covariate X and a binary response Y , the marginal model is given by:

$$\begin{aligned}\Pr(Y_{ij} = 1) &= E(Y_{ij} | x_{ij}) = \mu_{ij} \\ \text{logit}(\mu_{ij}) &= \log\left(\frac{\mu_{ij}}{1 - \mu_{ij}}\right) = \beta_0 + \beta_1 x_{ij} \quad (\text{log odds}) \\ \text{Var}(Y_{ij}) &= \mu_{ij}(1 - \mu_{ij}) \\ \text{Cor}(Y_{ij}, Y_{ik}) &= \rho \quad \text{for example}\end{aligned}$$

ICHS example: For the j th visit of the i th individual

$$\begin{aligned}x_{ij} &= \begin{cases} 1 & \text{Vitamin A deficient} \\ 0 & \text{otherwise} \end{cases} \\ y_{ij} &= \begin{cases} 1 & \text{respiratory infection} \\ 0 & \text{otherwise} \end{cases}\end{aligned}$$

- Odds of being infected ($y_{ij} = 1$) given $x_{ij} = 0$ is e^{β_0}
- Odds of being infected ($y_{ij} = 1$) given $x_{ij} = 1$ is $e^{\beta_0 + \beta_1}$
- Odds ratio (Vitamin A deficient vs Vitamin A replete) is e^{β_1} .
- β_0 and β_1 are population-average parameters.
- Interpretations of β_0 and β_1 remain the same regardless of n_i and of magnitude of within-cluster dependence.

Random Effects Model

For binary data, a random intercept model can be written as

$$\begin{aligned} \text{logit Pr}(Y_{ij} = 1 | u_i) &= \beta_0^* + u_i + \beta_1^* x_{ij} \\ u_i &\stackrel{\text{iid}}{\sim} \mathcal{N}(0, v^2) \end{aligned}$$

ICHS example

- Probability of infection for child i when vitamin A replete ($x_{ij} = 0$) is

$$\Pr(Y_{ij} = 1 | u_i, x_{ij} = 0) = \frac{e^{\beta_0^* + u_i}}{1 + e^{\beta_0^* + u_i}},$$

with corresponding odds of infection:

$$e^{\beta_0^* + u_i}$$

- If that child becomes vitamin A deficient ($x_{ij} = 1$), the odds of respiratory infection is

$$e^{\beta_0^* + u_i + \beta_1^*},$$

giving an odds ratio of infection (vitamin A deficient vs vitamin A replete)

$$e^{\beta_1^*}.$$

- $e^{\beta_1^*}$ is the odds ratio of respiratory infection for vitamin A deficiency of an “average” child with $u_i = 0$.
- How are (β_0^*, β_1^*) and (β_0, β_1) related?

The marginal probability of infection is

$$\begin{aligned} \Pr(Y_{ij} = 1 | x_{ij}) &= \int_{-\infty}^{\infty} \Pr(Y_{ij} = 1 | u_i, x_{ij}) f(u_i; v^2) du_i \\ &= \int_{-\infty}^{\infty} \frac{e^{\beta_0^* + u_i + \beta_1^* x_{ij}}}{1 + e^{\beta_0^* + u_i + \beta_1^* x_{ij}}} \frac{e^{-\frac{u_i^2}{2v^2}}}{\sqrt{2\pi v^2}} du_i \\ &= \frac{e^{\beta_0 + \beta_1 x_{ij}}}{1 + e^{\beta_0 + \beta_1 x_{ij}}} \quad \text{under marginal model} \end{aligned}$$

The relationship between (β_0^*, β_1^*) and (β_0, β_1) is complicated. Some theoretical results (for binary data with logit link):

- (Neuhaus et al., 1991) If $\text{Var}(u_i) = v^2 > 0$,

$$|\beta_k| \leq |\beta_k^*|, \quad k = 1, \dots, p$$

where the equality holds only when $\beta^* = 0$. Moreover, the discrepancy between β and β^* increases with v^2 .

- (Zeger et al., 1988) If $u_i \sim \mathcal{N}(0, v^2)$,

$$\beta \approx \frac{1}{\sqrt{0.346v^2 + 1}}\beta^*.$$

As v^2 increases, β is closer to zero.

ICHS example

Assume $\beta_0^* = -2.0$, $\beta_1^* = 0.4$ and $\text{Var}(u_i) = v^2 = 2.0$.

Value of u_i	$\Pr(1 u_i, 0)$	$\Pr(1 u_i, 1)$	odds ratio
u_i at 2.5% = $-2\sqrt{2}$	0.084	0.13	1.49
u_i at 50% = 0	0.12	0.17	1.49
u_i at 97.5% = $2\sqrt{2}$	0.68	0.76	1.49

- Absolute risk of infection increases more for children who have higher propensity for infection ($u_i > 0$).
- Marginal model:

$$\Pr(y_{ij} = 1 | x_{ij} = 0) = \int_{-\infty}^{\infty} \frac{e^{-2+u_i}}{1 + e^{-2+u_i}} \frac{e^{-u_i^2/4}}{\sqrt{4\pi}} du_i = 0.18$$

$$\Pr(y_{ij} = 1 | x_{ij} = 1) = \int_{-\infty}^{\infty} \frac{e^{-2+u_i+0.4}}{1 + e^{-2+u_i+0.4}} \frac{e^{-u_i^2/4}}{\sqrt{4\pi}} du_i = 0.23$$

The odds ratio for vitamin A deficiency is

$$e^{\beta_1} = \frac{0.23/(1 - 0.23)}{0.18/(1 - 0.18)} \approx 1.36.$$

So $\beta_1 = \log(1.36) = 0.31$.

- See Figures 7.2 and 7.3 in DHLZ

Transition Models

If Y_{ij} depends only on the last observation (Markov model, discrete autoregressive model with lapse 1)

$$\text{logit Pr}(y_{ij} = 1 \mid x_{ij}, H_{ij}) = x_{ij}\beta^{**} + \alpha y_{i,j-1}$$

- $\exp(\beta^{**})$ is the odds ratio for vitamin A deficient vs. replete among children who were free of infection at last visit (or same status of infection at last visit)
- $\exp(\alpha)$ is the odds ratio for children who did and did not have infection at last visit.

The Markov chain has transition probability matrix

		$y_{ij} =$	
		0	1
$y_{i,j-1} =$	0	$\frac{1}{1+\exp(x_{ij}\beta^{**})}$	$1 - \frac{1}{1+\exp(x_{ij}\beta^{**})}$
	1	$\frac{1}{1+\exp(x_{ij}\beta^{**}+\alpha)}$	$1 - \frac{1}{1+\exp(x_{ij}\beta^{**}+\alpha)}$

- Transition probability matrices vary with x_{ij} and therefore across individuals.
- The initial distribution (for stationary chain, or otherwise) can be difficult to specify.

Further Reading

- Chapter 7 of DHLZ.