

1 Conditional linear odds model using Newton-Raphson Algorithm

1.1 Basic linear odds model (no covariates)

Basic linear odds model means the regression model does not control the covariates, which is written as:

$$O_i = \text{odds}_i = \exp(\beta_0)(1 + \beta_1 G_i + \beta_2 E_i + \beta_3 G_i E_i)$$

where

$$O_i = \frac{p_i}{1 - p_i} \Leftrightarrow p_i = \frac{O_i}{1 + O_i}, \quad p_i = p(Y_i = 1 \mid G_i = g, E_i = e)$$

Then the likelihood function of (y_1, \dots, y_n) can be written as:

$$\mathcal{L}(\beta) = \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1 - y_i}$$

and the log-likelihood:

$$\ell(\beta) = \sum_{i=1}^n y_i \log(p_i) + (1 - y_i) \log(1 - p_i)$$

It can be also written in terms of beta explicitly:

$$\ell(\beta) = \sum_{i=1}^n y_i \log \left(\frac{\exp(\beta_0) z_i}{1 + \exp(\beta_0) z_i} \right) + (1 - y_i) \log \left(\frac{1}{1 + \exp(\beta_0) z_i} \right)$$

where

$$p_i = \frac{\exp(\beta_0) z_i}{1 + \exp(\beta_0) z_i}, \quad z_i = 1 + \beta_1 G_i + \beta_2 E_i + \beta_3 G_i E_i$$

Score function:

Define the score vector:

$$S(\beta)_j = \frac{\partial \ell(\beta)}{\partial \beta_j}$$

It can be further expanded as

$$\begin{aligned} \frac{\partial \ell}{\partial \beta_j} &= \sum_{i=1}^n \left(\frac{y_i}{p_i} - \frac{1 - y_i}{1 - p_i} \right) \frac{\partial p_i}{\partial \beta_j} \\ &= \sum_{i=1}^n \frac{y_i - p_i}{p_i(1 - p_i)} \frac{\partial p_i}{\partial O_i} \frac{\partial O_i}{\partial \beta_j} \\ &= \sum_{i=1}^n \frac{r_i}{O_i} \frac{\partial O_i}{\partial \beta_j} \end{aligned}$$

where

$$r_i = y_i - p_i, \quad \frac{dp_i}{dO_i} = \frac{1}{(1 + O_i)^2}$$

For each β_j , we have:

$$\frac{\partial O_i}{\partial \beta_j} = \begin{cases} O_i & j = 0 \\ \exp(\beta_0) G_i & j = 1 \\ \exp(\beta_0) E_i & j = 2 \\ \exp(\beta_0) G_i E_i & j = 3 \end{cases}$$

Consequently

$$S(\beta)_j = \begin{cases} \sum_{i=1}^n r_i & j = 0 \\ \sum_{i=1}^n \frac{r_i G_i}{z_i} & j = 1 \\ \sum_{i=1}^n \frac{r_i E_i}{z_i} & j = 2 \\ \sum_{i=1}^n \frac{r_i G_i E_i}{z_i} & j = 3 \end{cases}$$

Hence

$$S(\beta) = \left(\sum_{i=1}^n r_i, \quad \sum_{i=1}^n \frac{r_i G_i}{z_i}, \quad \sum_{i=1}^n \frac{r_i E_i}{z_i}, \quad \sum_{i=1}^n \frac{r_i G_i E_i}{z_i} \right)^\top$$

Observed Information Matrix

The information is defined as

$$I(\beta)_{ij} = -\frac{\partial}{\partial \beta_j} S(\beta)_i$$

Denote w_i as

$$w_i = \frac{O_i}{(1 + O_i)^2}$$

By derivations, the final $\mathcal{I}(\beta)$ is written as

$$I(\beta) = \begin{pmatrix} \sum w_i & \sum w_i \frac{G_i}{z_i} & \sum w_i \frac{E_i}{z_i} & \sum w_i \frac{G_i E_i}{z_i} \\ \sum w_i \frac{G_i}{z_i} & \sum (w_i + r_i) \frac{G_i^2}{z_i^2} & \sum (w_i + r_i) \frac{G_i E_i}{z_i^2} & \sum (w_i + r_i) \frac{G_i^2 E_i}{z_i^2} \\ \sum w_i \frac{E_i}{z_i} & \sum (w_i + r_i) \frac{G_i E_i}{z_i^2} & \sum (w_i + r_i) \frac{E_i^2}{z_i^2} & \sum (w_i + r_i) \frac{G_i E_i^2}{z_i^2} \\ \sum w_i \frac{G_i E_i}{z_i} & \sum (w_i + r_i) \frac{G_i^2 E_i}{z_i^2} & \sum (w_i + r_i) \frac{G_i E_i^2}{z_i^2} & \sum (w_i + r_i) \frac{(G_i E_i)^2}{z_i^2} \end{pmatrix}$$

The Newton-Raphson Algorithm can be therefore written as

$$\hat{\beta}^{(r+1)} = \hat{\beta}^{(r)} + I \left(\hat{\beta}^{(r)} \right)^{-1} S \left(\hat{\beta}^{(r)} \right)$$

where \mathcal{I} and S are defined above.

1.2 Linear odds model with one covariate (conditional linear odds regression)

The linear odds model now changes to

$$O_i = \exp(\beta_0 + \gamma_1 X_{1i}) (1 + \beta_1 G_i + \beta_2 E_i + \beta_3 G_i E_i)$$

Then the log-likelihood becomes:

$$\ell(\beta) = \sum_{i=1}^n [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)]$$

where

$$p_i = \frac{\exp(\beta_0 + \gamma_1 X_{1i}) z_i}{1 + \exp(\beta_0 + \gamma_1 X_{1i}) z_i}, \quad z_i = 1 + \beta_1 G_i + \beta_2 E_i + \beta_3 G_i E_i$$

Score Function

Let $\theta = (\beta, \gamma_1)$, the score vecot can be written as

$$S(\theta) = \begin{pmatrix} \frac{\partial \ell}{\partial \beta} \\ \frac{\partial \ell}{\partial \gamma_1} \end{pmatrix}$$

Similar as before, the derivatives wrt β_j can be written as

$$\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^n \left(\frac{y_i}{p_i} - \frac{1 - y_i}{1 - p_i} \right) \frac{\partial p_i}{\partial \beta_j} = \sum_{i=1}^n \frac{r_i}{O_i} \cdot \frac{\partial O_i}{\partial \beta_j}$$

where $r_i = y_i - p_i$. We can follow the exactly the same procedure as previous, and the scores for β are

$$S(\beta)_j = \begin{cases} \sum_{i=1}^n r_i & j = 0 \\ \sum_{i=1}^n \frac{r_i G_i}{z_i} & j = 1 \\ \sum_{i=1}^n \frac{r_i E_i}{z_i} & j = 2 \\ \sum_{i=1}^n \frac{r_i G_i E_i}{z_i} & j = 3 \end{cases}$$

For γ_1 , its score can be written as

$$\frac{\partial \ell}{\partial \gamma_1} = \sum_{i=1}^n \left(\frac{y_i}{p_i} - \frac{1 - y_i}{1 - p_i} \right) \cdot \frac{\partial p_i}{\partial \gamma} = \sum_{i=1}^n \frac{r_i}{O_i} \cdot \frac{\partial O_i}{\partial \gamma}$$

Since $\frac{\partial O_i}{\partial \gamma_1} = O_i X_{1i}$, the score for γ_1 is

$$S(\beta)_4 = \sum_{i=1}^n r_i X_{1i}$$

Information Matrix

The derivations for $I_{00}, I_{01}, \dots, I_{33}$ are same as previous. In addition, the derivations for I_{40} to I_{44} follows a similar way. For example

$$\begin{aligned} I_{40} = I_{04} &= -\frac{\partial}{\partial \gamma} S(\beta_0) \\ &= -\sum_{i=1}^n \left(-\frac{\partial p_i}{\partial \gamma} \right) \\ &= \sum_{i=1}^n \frac{1}{(1 + O_i)^2} \cdot \frac{\partial O_i}{\partial \gamma} \\ &= \sum_{i=1}^n w_i X_i \end{aligned}$$

where

$$w_i = \frac{O_i}{(1 + O_i)^2}$$

Then the final information matrix is written as

$$I(\boldsymbol{\theta}) = \begin{pmatrix} \sum w_i & \sum w_i \frac{G_i}{z_i} & \sum w_i \frac{E_i}{z_i} & \sum w_i \frac{G_i E_i}{z_i} & \sum w_i X_{1i} \\ \sum w_i \frac{G_i}{z_i} & \sum (w_i + r_i) \frac{G_i^2}{z_i^2} & \sum (w_i + r_i) \frac{G_i E_i}{z_i^2} & \sum (w_i + r_i) \frac{G_i^2 E_i}{z_i^2} & \sum w_i X_{1i} \frac{G_i}{z_i} \\ \sum w_i \frac{E_i}{z_i} & \sum (w_i + r_i) \frac{G_i E_i}{z_i^2} & \sum (w_i + r_i) \frac{E_i^2}{z_i^2} & \sum (w_i + r_i) \frac{G_i E_i^2}{z_i^2} & \sum w_i X_{1i} \frac{E_i}{z_i} \\ \sum w_i \frac{G_i E_i}{z_i} & \sum (w_i + r_i) \frac{G_i^2 E_i}{z_i^2} & \sum (w_i + r_i) \frac{G_i E_i^2}{z_i^2} & \sum (w_i + r_i) \frac{(G_i E_i)^2}{z_i^2} & \sum w_i X_{1i} \frac{G_i E_i}{z_i} \\ \sum w_i X_{1i} & \sum w_i X_{1i} \frac{G_i}{z_i} & \sum w_i X_{1i} \frac{E_i}{z_i} & \sum w_i X_{1i} \frac{G_i E_i}{z_i} & \sum w_i X_{1i}^2 \end{pmatrix}$$

Similarly, the Newton-Raphson Algorithm can be therefore written as

$$\hat{\boldsymbol{\beta}}^{(r+1)} = \hat{\boldsymbol{\beta}}^{(r)} + I \left(\hat{\boldsymbol{\beta}}^{(r)} \right)^{-1} S \left(\hat{\boldsymbol{\beta}}^{(r)} \right)$$

where \mathcal{I} and S are defined above.

1.3 Linear odds model with two covariates

The linear odds model now changes to

$$O_i = \exp(\beta_0 + \gamma_1 X_{1i} + \gamma_2 X_{2i}) (1 + \beta_1 G_i + \beta_2 E_i + \beta_3 G_i E_i)$$

Then the log-likelihood becomes:

$$\ell(\beta) = \sum_{i=1}^n [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)]$$

where

$$p_i = \frac{\exp(\beta_0 + \gamma_1 X_{1i} + \gamma_2 X_{2i}) z_i}{1 + \exp(\beta_0 + \gamma_1 X_{1i} + \gamma_2 X_{2i}) z_i}, \quad z_i = 1 + \beta_1 G_i + \beta_2 E_i + \beta_3 G_i E_i$$

Score Function

The scores for the β are the same as previous. For γ_1 and γ_2 , their scores are

$$\begin{aligned} \frac{\partial O_i}{\partial \gamma_1} &= O_i X_{i1} \Rightarrow S(\gamma_1) = \sum_{i=1}^n r_i X_{i1} \\ \frac{\partial O_i}{\partial \gamma_2} &= O_i X_{i2} \Rightarrow S(\gamma_2) = \sum_{i=1}^n r_i X_{i2} \end{aligned}$$

where

$$r_i = y_i - p_i, \quad z_i = 1 + \beta_1 G_i + \beta_2 E_i + \beta_3 G_i E_i$$

Information Matrix

The cross term between the two covariates are

$$I_{45} = I_{54} = \frac{\partial}{\partial \gamma_2} S(\gamma_1) = \sum_{i=1}^n X_{1i} X_{2i} w_i$$

and the rest terms are same as before.

Then the final information matrix is written as

$$I(\theta) = \begin{pmatrix} \sum w_i & \sum w_i \frac{G_i}{z_i} & \sum w_i \frac{E_i}{z_i} & \sum w_i \frac{G_i E_i}{z_i} & \sum w_i X_{i1} & \sum w_i X_{i2} \\ \sum w_i \frac{G_i}{z_i} & \sum (w_i + r_i) \frac{G_i^2}{z_i^2} & \sum (w_i + r_i) \frac{G_i E_i}{z_i^2} & \sum (w_i + r_i) \frac{G_i^2 E_i}{z_i^2} & \sum w_i X_{i1} \frac{G_i}{z_i} & \sum w_i X_{i2} \frac{G_i}{z_i} \\ \sum w_i \frac{E_i}{z_i} & \sum (w_i + r_i) \frac{G_i E_i}{z_i^2} & \sum (w_i + r_i) \frac{E_i^2}{z_i^2} & \sum (w_i + r_i) \frac{G_i E_i^2}{z_i^2} & \sum w_i X_{i1} \frac{E_i}{z_i} & \sum w_i X_{i2} \frac{E_i}{z_i} \\ \sum w_i \frac{G_i E_i}{z_i} & \sum (w_i + r_i) \frac{G_i^2 E_i}{z_i^2} & \sum (w_i + r_i) \frac{G_i E_i^2}{z_i^2} & \sum (w_i + r_i) \frac{(G_i E_i)^2}{z_i^2} & \sum w_i X_{i1} \frac{G_i E_i}{z_i} & \sum w_i X_{i2} \frac{G_i E_i}{z_i} \\ \sum w_i X_{i1} & \sum w_i X_{i1} \frac{G_i}{z_i} & \sum w_i X_{i1} \frac{E_i}{z_i} & \sum w_i X_{i1} \frac{G_i E_i}{z_i} & \sum w_i X_{i1}^2 & \sum w_i X_{i1} X_{i2} \\ \sum w_i X_{i2} & \sum w_i X_{i2} \frac{G_i}{z_i} & \sum w_i X_{i2} \frac{E_i}{z_i} & \sum w_i X_{i2} \frac{G_i E_i}{z_i} & \sum w_i X_{i1} X_{i2} & \sum w_i X_{i2}^2 \end{pmatrix}$$

1.4 Linear oddds model with more than two covariates (generalized version)

We can generalize the previous idea to linear odds model with k covariates such that $k > 2$. The model is written as

$$O_i = \exp \left(\beta_0 + \sum_{m=1}^k \gamma_m X_{mi} \right) (1 + \beta_1 G_i + \beta_2 E_i + \beta_3 G_i E_i)$$

Denote $\boldsymbol{\theta} = (\beta_0, \dots, \beta_3, \gamma_1, \dots, \gamma_k)$. Also denote \mathbf{d}_i such that

$$\mathbf{d}_i := \begin{pmatrix} \mathbf{v}_i^\beta \\ \mathbf{x}_i \end{pmatrix} \in \mathbb{R}^d$$

. where $\mathbf{v}_i^\beta = (1, a_i, b_i, c_i)^\top$ such that $a_i := \frac{G_i}{z_i}, b_i := \frac{E_i}{z_i}, c_i := \frac{G_i E_i}{z_i}$ and $\mathbf{x}_i := (X_{1i}, \dots, X_{ki})^\top$. The score vector can be written as

$$\mathbf{S}(\boldsymbol{\theta}) = \sum_{i=1}^n r_i \mathbf{d}_i$$

For the information matrix, $\mathcal{I}(\boldsymbol{\theta})$ can be written as block block structure with dimension $(4+k) \times (4+k)$ such that

$$I(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{I}_{\beta\beta} & \mathbf{I}_{\beta\gamma} \\ \mathbf{I}_{\beta\gamma}^\top & \mathbf{I}_{\gamma\gamma} \end{pmatrix} \quad (4+k) \times (4+k)$$

where

$$\mathbf{I}_{\beta\beta} = \begin{pmatrix} \sum w_i & \sum w_i \frac{G_i}{z_i} & \sum w_i \frac{E_i}{z_i} & \sum w_i \frac{G_i E_i}{z_i} \\ \sum w_i \frac{G_i}{z_i} & \sum (w_i + r_i) \frac{G_i^2}{z_i^2} & \sum (w_i + r_i) \frac{G_i E_i}{z_i^2} & \sum (w_i + r_i) \frac{G_i^2 E_i}{z_i^2} \\ \sum w_i \frac{E_i}{z_i} & \sum (w_i + r_i) \frac{G_i E_i}{z_i^2} & \sum (w_i + r_i) \frac{E_i^2}{z_i^2} & \sum (w_i + r_i) \frac{G_i E_i^2}{z_i^2} \\ \sum w_i \frac{G_i E_i}{z_i} & \sum (w_i + r_i) \frac{G_i^2 E_i}{z_i^2} & \sum (w_i + r_i) \frac{G_i E_i^2}{z_i^2} & \sum (w_i + r_i) \frac{(G_i E_i)^2}{z_i^2} \end{pmatrix}$$

and

$$\begin{aligned} \mathbf{I}_{\beta\gamma} &= \sum_i w_i \mathbf{v}_i^\beta \mathbf{x}_i^\top \\ &= \begin{pmatrix} \sum w_i X_{1i} & \cdots & \sum w_i X_{ki} \\ \sum w_i a_i X_{1i} & \cdots & \sum w_i a_i X_{ki} \\ \sum w_i b_i X_{1i} & \cdots & \sum w_i b_i X_{ki} \\ \sum w_i c_i X_{1i} & \cdots & \sum w_i c_i X_{ki} \end{pmatrix} \end{aligned}$$

For $\mathbf{I}_{\gamma\gamma}$, it is expressed as

$$\begin{aligned} \mathbf{I}_{\gamma\gamma} &= \sum_i w_i \mathbf{x}_i \mathbf{x}_i^\top \\ &= \begin{pmatrix} \sum w_i X_{1i}^2 & \cdots & \sum w_i X_{1i} X_{ki} \\ \vdots & \ddots & \vdots \\ \sum w_i X_{ki} X_{1i} & \cdots & \sum w_i X_{ki}^2 \end{pmatrix} \end{aligned}$$

which looks like covariance.

The way we use linear odds regression because it has better performance when the outcome is NOT rare. The difficulty of causal interaction is that we can not easily access the RERI in additive scale when the outcome is not rare. Linear odds model is designed to solve this problem.