Chapter 8

Random Effects Models for Repeated Binary Data

The models and methods for repeated binary data which were considered in Chapter 7 are most appropriate when the data are balanced, that is, there are n common occasions of measurement and imbalance ($n_i \neq n$ for some i) arises because of missing observations. When the unequal n_i arise because of inherently unbalanced data or because of clustered designs, the most natural approach is to consider extending the LMM using random effects to the GLM setting.

By analogy to the linear case, we assume each subject has a vector of subject-specific effects, b_i, and we add $Z_i b_i$ to the linear predictor $X_i\beta$. Letting Y_i denote the $n_i \times 1$ vector of binary outcomes, we have

$$E(Y_i \mid b_i, X_i) = \mu_i^* = g(X_i\beta^* + Z_i b_i) \tag{8.1}$$

where

$$\ell\,(\mu_i^*) = X_i\beta^* + Z_i b_i, \tag{8.2}$$

ℓ is the link function, and g the inverse link function. As before, we assume that $E(b_i) = 0$ and $\text{var}(b_i) = D$. Generally, we also assume that given b_i, the Y_{ij}'s are independent.

We use the μ_i^*, β^* notation to emphasize that μ_i^* and β^* are conditional and not marginal parameters. Recall that for μ_i, β defined in Chapters 6 and 7, we assume that

$$E(Y_i \mid X_i) = \mu_i = g(X_i\beta).$$

But here we have

$$E(Y_i \mid X_i, b_i) = \mu_i^* = g(X_i\beta^* + Z_i b_i)$$