Longitudianl Data Analysis

3. Linear Models for Correlated Data

To develop a general linear model framework for longitudinal data, in which the inference we make about the parameters of interest (β) *recoginze the likely correlation structure of the data*. There are two ways to archieve this:

1. To build explicit parametric models of the covariance struture (e.g. CS/exchangable, AR(1), etc.)
2. To use methods of inference which are robust to misspecification of the covariance structure.[1][2]

1. Random effect model (conditional model)

$$Y_{ij} = U_i + \beta_0 + \beta_1 t_j + \varepsilon_{ij}$$

where

$$U_i \sim N(0, v^2), \text{ variance between units (clusters)}$$

$$\varepsilon_{ij} \sim N(0, \tau^2), \text{ variation within units}$$

$$\rho = \frac{v^2}{v^2 + \tau^2}, \text{ proportion of total variance due to between units variation}$$

- $var(y_{ij}|U_i) = \tau^2, cov(y_{ij}, y_{ij'}|U_i) = 0$
- Ex: *R.E. (repeated measures) ANOVA*:

  we can treat the group (condition) variable as the orignial "time" variable, i.e. the repeated meaures are observed on different conditions (which cause the groups effect that we are interested in).

  $$y_{ij} = \mu_j + U_i + \varepsilon_{ij}$$

  where

  $$\mu_j, j = 1, \cdots, n \text{ the group effects}$$

  $$U_i \sim N(0, v^2), \text{ the between subjects random effect}$$

  $$\varepsilon \sim N(0, \tau^2)$$

  Total sum of squares= Between group sum of squares+within group sum of squares

[1] The covariance structure is usually the nuisance parameter

[2] Two categories of robustness: 1. w.r.t. the outliers; 2. w.r.t. model assumptions

$$cov(y_{ij}, y_{i'j'}) = \begin{cases} var(y_{ij}) = \tau^2 + v^2, & \text{if } i = i', j = j' \\ cov(y_{ij}, y_{ij'}) = v^2, & \text{if } i = i', j \neq j' \\ 0\ , & \text{if } i \neq i', j \neq j' \end{cases}$$

$$corr(y_{ij}, y_{ij'}) = \frac{cov(y_{ij}, y_{ij'})}{\sqrt{var(y_{ij})var(y_{ij'})}} = \frac{v^2}{\tau^2 + v^2} = \rho \text{ (within cluster correlation)}$$

- ***heteogeneity***: variantion between groups implies (relative) similarity/correlation within gorups. If $v^2 >> \tau^2 \Rightarrow \rho \to 1 \Rightarrow$ significant difference among groups.

2. Marginal model (population average)

1. With exchangable (CS) variance-covariance structure (balanced data)

$$Y_{ij} = \beta_0 + \beta_1 t_j + \varepsilon_{ij}, i = 1, /cdots, m; j = 1, \cdots, n$$

$$Cov(Y_i) = \sigma^2 V_0$$

where

$$V_0 = \rho J + (1 - \rho) I$$

$$\rho = corr(y_{ij}, y_{ik}), \text{ for } i \neq k$$

   - marginal model with CS variance struture is equivalent with random effect (intercept) model: the same $E(Y_i)$ and $cov(Y_i)$ structures, thus will return the same estimate of the parameters.

2. Exponential correlation / autoregressive model

   - Correlation b/t a pair of measurements of the same subject decays towards 0 as the time separation b/t the measurement increases, i.e.
     - for unequally-spaced data

$$cov(y_{ij}, y_{ik}) = [V_0]_{ij} = v_{ij} = \sigma^2 e^{-\phi |t_j - t_k|}$$

     - for equally-spaced data

$$t_{j-1} - t_j = d$$

     for all j, then $v_{jk} = \sigma^2 \rho$, where $\rho = \exp(-\phi d)$ is the correlation b/t successive observations on the same subject.

- In marginal models, we model the mean and the covariance structrue separately:

$$E[Y_i] = \mu = X_i \beta$$

$$Cov[Y_i] = CS/AR(1)/...$$

- Three basic elements of correlation structure:
  - Random effects
  - Autocorrelation or serial dependence
  - Noise, measurement error

- Modelling the correlation in longitudianl data is important to be able to obtain correct inference on βs. Incorporating correlation into estimation of regression models is achieved via **weighted least squares**.