Longitudianl Data Analysis

Cross-sectional v.s. Longitudianl study

Cross-sectional (CS) study can be reduced from the longitudinal study (LS) if the number of measures per subject is equal to one, ie $n_i = 1$.

$$Y_{i1} = \beta_c x_{i1} + \varepsilon_{i1}, i = 1, \cdots, m$$

- β_c represents the diff in average Y across two sub-populations which differ by one unit in x.

With repeated measures, above model becomes LDA model:

$$Y_{ij} = \beta_c x_{i1} + \beta_L (x_{ij} - x_{i1}) + \varepsilon_{ij}, i = 1, \cdots, m; j = 1, \cdots, n_i$$

- Notice that β_L represents the expected change in Y over time per one unit change in x, w.r.t its baseline value:

$$(Y_{ij} - Y_{i1}) = \beta_L (x_{ij} - x_{i1}) + \varepsilon_{ij} - \varepsilon_{i1}$$

- When $n = 1$, above two models are identical.
- It's more common that β_C and β_L have the same sign. However, it may exist that they have opposite sign.
- In CS the basis is a comparison of individuals with a particular value of x to others with a different value
- In LDA each person is his/her own control. β_L is estimated by comparing a person's response at two times assuming that x changes over time.
- Estimation of β_C is confouned by unmeasured individual characteristic; while estimation of β_L is less likely to be affected by unmeasured confounding. (since, if the confounders are not time-varing, they will be cancelled out when doing model (2)-(1)).