STAT 902: Advanced Data Analysis

Whiteboard Lecture

Longitudinal Data Analysis:

Whiteboard Interlude II

We now describe the essential properties of the *random intercept model*

$$y_{ij} = U_i + \beta_0 + \beta_1 x_{ij} + \varepsilon_{ij}$$

$$U_i \overset{\text{ind.}}{\sim} N(0, \sigma_U^2), \qquad \varepsilon_{ij} \overset{\text{ind.}}{\sim} N(0, \sigma_\varepsilon^2)$$

and the U_i and ε_{ij} are independent of each other.

For $j \neq j'$, Cov$(y_{ij}, y_{ij'})$ is the covariance between different measurements on the same subject:

$$\begin{aligned}
\text{Cov}(y_{ij}, y_{ij'}) &= \text{Cov}(U_i + \beta_0 + \beta_1 x_{ij} + \varepsilon_{ij}, \\
&\qquad U_i + \beta_0 + \beta_1 x_{ij'} + \varepsilon_{ij'}) \\
&= \text{Cov}(U_i + \varepsilon_{ij}, U_i + \varepsilon_{ij'}) \\
&= \text{Cov}(U_i, U_i) + \text{Cov}(U_i, \varepsilon_{ij'}) + \text{Cov}(U_i, \varepsilon_{ij}) \\
&\quad + \text{Cov}(\varepsilon_{ij}, \varepsilon_{ij'}) \\
&= \sigma_U^2 + 0 + 0 + 0 \\
&= \sigma_U^2
\end{aligned}$$

For $j = j'$, Cov$(y_{ij}, y_{ij'})$ = Var(y_{ij}) is the variance of the jth measurement on subject i, and has expression:

$$\begin{aligned}
\text{Cov}(y_{ij}, y_{ij}) &= \text{Var}(U_i + \beta_0 + \beta_1 x_{ij} + \varepsilon_{ij}) \\
&= \text{Var}(U_i + \varepsilon_{ij}) \\
&= \text{Var}(U_i) + \text{Var}(\varepsilon_{ij}) \\
&= \sigma_U^2 + \sigma_\varepsilon^2
\end{aligned}$$

For $i \neq i'$ we get

$$\text{Cov}(y_{ij}, y_{i'j'}) = 0.$$

This says that observations of different individuals are uncorrelated (e.g. Anne's blood pressure is not correlated with Bill's blood pressure).

Consider the sample sizes corresponding to the example:

$$m = 3, \qquad n_1 = 2, \qquad n_2 = 3, \qquad n_3 = 2.$$

The covariance matrix of

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{31} \\ y_{32} \end{bmatrix}$$

is

$$\left[\begin{array}{cc|ccc|cc}
\sigma_U^2+\sigma_\varepsilon^2 & \sigma_U^2 & 0 & 0 & 0 & 0 & 0 \\
\sigma_U^2 & \sigma_U^2+\sigma_\varepsilon^2 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & \sigma_U^2+\sigma_\varepsilon^2 & \sigma_U^2 & \sigma_U^2 & 0 & 0 \\
0 & 0 & \sigma_U^2 & \sigma_U^2+\sigma_\varepsilon^2 & \sigma_U^2 & 0 & 0 \\
0 & 0 & \sigma_U^2 & \sigma_U^2 & \sigma_U^2+\sigma_\varepsilon^2 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & \sigma_U^2+\sigma_\varepsilon^2 & \sigma_U^2 \\
0 & 0 & 0 & 0 & 0 & \sigma_U^2 & \sigma_U^2+\sigma_\varepsilon^2
\end{array}\right]$$

The correlation matrix is then

$$\left[\begin{array}{cc|ccc|cc}
1 & \rho & 0 & 0 & 0 & 0 & 0 \\
\rho & 1 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 1 & \rho & \rho & 0 & 0 \\
0 & 0 & \rho & 1 & \rho & 0 & 0 \\
0 & 0 & \rho & \rho & 1 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 1 & \rho \\
0 & 0 & 0 & 0 & 0 & \rho & 1
\end{array}\right] \quad \text{where} \quad \rho = \frac{\sigma_U^2}{\sigma_U^2+\sigma_\varepsilon^2}.$$

Note the blocking structure corresponding to within-subject correlation. The first block is for subject 1 (Anne), the second is for subject 2 (Bill) and the third is for subject 3 (Clare).

Remarks

1. The random intercept U_i invokes correlation between measurements on same subject.
2. A shortcoming of the random intercept model is that the correlation is the same for each subject; e.g. Anne's ρ is the same as Bill's ρ.
3. Another shortcoming is that the within-subject correlation is constant over time; e.g. the correlation between Anne's blood pressure measurements 2 days apart is the same as those taken 10 days apart.

♦ ♦ ♦