NTNU
Norwegian University of
Science and Technology

# KLMED8008 Analysis of repeated measurements

Øyvind Salvesen

# Multilevel nested data

Two-level nested data:

- patients in hospitals
- siblings in families
- twins in twin-pairs

Three-level nested data:

- patients in hospitals in countries
- siblings in families in districts
- occasions in twins in twin-pairs

The data is nested in the sense that a lower level unit can only belong to one higher level unit.

# Random intercept models

$$y_{i,j} = (\beta_1 + \zeta_j) + \beta_2 x_{i,j} + \epsilon_{i,j}$$

$$\zeta_j \sim N(0, \psi)$$

$$\epsilon_{i,j} \sim N(0, \theta)$$

Variables in $\{\zeta_j : j = 1, \ldots, J\} \cup \{\epsilon_{i,j} : j = 1, \ldots, J, i = 1, \ldots, n_j\}$ are assumed independent, thus the distribution of observables $\{y_{i,j}\}$

$$f(\{y_{i,j}\}; \beta_1, \beta_2, \psi, \theta)$$

is specified.

# Random coefficient models

$$y_{i,j} = (\beta_1 + \zeta_{1,j}) + (\beta_2 + \zeta_{2,j})x_{i,j} + \epsilon_{i,j}$$

The pair $\zeta_{\mathbf{j}} = (\zeta_{1,j}, \zeta_{2,j})$ is assumed bivariate normal with mean zero and covariance matrix

$$\Psi = \begin{pmatrix} \psi_{1,1} & \sqrt{\psi_{1,1}}\sqrt{\psi_{2,2}}\rho \\ & \psi_{2,2} \end{pmatrix}$$

$$\epsilon_{i,j} \sim N(0, \theta)$$

Variables in $\{\zeta_{\mathbf{j}}\} \cup \{\epsilon_{i,j}\}$ are assumed independent, thus the distribution of observables $\{y_{i,j}\}$

$$f(\{y_{i,j}\}; \beta_1, \beta_2, \psi_{1,1}, \psi_{2,2}, \rho, \theta)$$

is specified.

# Longitudinal data

Subjects repeatedly measured over time
Also called panel data, repeated measures, cross sectional time series
Examples:

- A randomized clinical trial recording scores and treatment over time.
- Students and their standardized test scores in six successive years.
- Hourly wages and explanatory variables (such as years of education) recorded over years.

In the special case that the repeated measurements given covariates are exchangeable (order does not matter), the data can also be viewed as two-level data with measurements nested in subjects.
For example, growth curve models is a special case of random coefficient models with covariates being functions (often polynomials or piecewice linear functions) of time.

# Longitudinal models

$$y_{i,j} = \beta x_{i,j} + \alpha_i + \zeta_j + \epsilon_{i,j}$$

$$\zeta_j \sim N(0, \psi)$$

$\epsilon_{\mathbf{j}} = (\epsilon_{1,j}, \epsilon_{2,j}, \ldots, \epsilon_{n_j})$ multivariate normal with mean zero and covariance matrix $\Sigma$. Variables in $\{\zeta_j\} \cup \{\epsilon_{\mathbf{j}}\}$ are assumed independent, thus the distribution of observables $\{y_{i,j}\}$

$$f(\{y_{i,j}\}; \beta, \{\alpha_i\}, \psi, \Sigma)$$

is specified.

Common specifications of $\Sigma$:

— Identity covariance structure:

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

— Autoregressive covariance structure of order 1:

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix}$$

— Exponential covariance structure:

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho^{t_2-t_1} & \rho^{t_3-t_1} \\ \rho^{t_2-t_1} & 1 & \rho^{t_3-t_2} \\ \rho^{t_3-t_1} & \rho^{t_3-t_2} & 1 \end{pmatrix}$$

— Toeplitz covariance structure:

$$\Sigma = \begin{pmatrix} a & b & c \\ b & a & b \\ c & b & a \end{pmatrix}$$

— Unstructured covariance:

$$\Sigma = \begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix}$$

## Example: A randomized clinical trial

$$y_{time,id} = \alpha_{time,treatment(id)} + \zeta_{id} + \epsilon_{time,id}$$

By randomization:

$$\alpha_{\text{baseline},1} = \alpha_{\text{baseline},2}$$

$$\zeta_{id} \sim N(0, \psi)$$

Covariance matrix of $\epsilon_{\mathbf{id}} = (\epsilon_{1,id}, \epsilon_{2,id}, \epsilon_{3,id})$:

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix}$$