Multilevel Models

Yongdai Kim

Seoul National University, Korea

0. **Outline**

1. Introduction
2. Multilevel analysis with Linear model
3. Multilevel analysis with Generalized linear model
4. Multilevel linear model vs multilevel GLM

1. Introduction

- What are "Multilevel Models"?
  - Multilevel models are statistical models for data displaying hierarchical structures.
  - Example of hierarchical structures
    * Nation → Region → County
    * Village → Household → Individual
    * Level 3 → Level 2 → Level 1

Example 1: Testis cancer mortality in the European country

- The data set consists of testis cancer mortality for males of all ages between 1971 and 1980 in 9 European countries.
- There are three levels
  - Level 1: county
  - Level 2: Region
  - Level 3: Country
- The objective of the study is to investigate the distribution of testis cancer mortality in relation to income and urban-rural status.

Example 2: Sleep pattern vs Cough (Repeated measures)

- Response variable: the percentage of the night spent awake.
- Explanatory variable: the total number of cough recorded during the night.
- 39 children were assessed on a number of nights varying from four to six.
- There are two levels:
  - Level 1: each night
  - Level 2: children

Example 3: UNICEF water sanitation intervention study (Cluster randomization)

- There are three levels:
  - Level 1: the number of diarrhea on every 2 months
  - Level 2: Children
  - Level 3: Village

Strengths of Multilevel Models

- Explicitly account for the interdependence of clustered units (where clustering may be spatial or temporal).
- Allow for the modeling of both average (fixed) effects and individual (random) effects.
- Facilitate thinking about and modeling context x person interactions.
- Permit inferences to be drawn to broader populations

Aim of the talk

- The aim of this talk is to review statistical models and inferential methods for multilevel data.
- First, statistical methodologies based on the linear models (i.e. continuous data) are reviewed,
- and methods based on generalized linear models (i.e binary or count data) are discussed.
- In particular, differences and difficulties of multilevel models based on the GLM compared to linear multilevel models are explained.

2. Multilevel analysis with linear models

- Consider the Sleep pattern vs Cough (SPC) data set (Level 1 - night, Level 2 - children).
- Let n be the number of clusters in Level 2 (i.e $n = 39$) and let n_i be the number of observations from the ith cluster in Level 2.
- X_{ij}: covariate from the jth observation in the ith cluster.
  - For SPC data, X_{ij} is the number of coughs at each night (Level 1 covariate).
  - We can use Level 2 covariates such as gender, age etc.
- Y_{ij}: response variable from the jth observation in the ith cluster.
  - One important feature of the multilevel model is that response variables from the same cluster are correlated.

<u>Variance component model</u>

- Model

  (a) $Y_{ij} = \beta_{0i} + \beta_1 X_{ij} + e_{ij}$

  (b) $\beta_{0i} = \beta_0 + u_{0i}$

  (c) $e_{ij} \sim N(0, \sigma^2), u_{0i} \sim N(0, \tau_{00}), Cov(e_{ij}, u_{0i}) = 0.$
- The effect of X to Y is (measured by β_1) equals for all clusters.
- The overall mean level of Y (after adjusting X) is β_0.
- But, the mean levels of Y (measured by β_{0i}) of clusters vary.
- The variance of β_{0i}, τ_{00} represents the degree of heterogeneity of clusters. Larger the variance is, more the mean levels of Y differ across the clusters.

- Observations in the same cluster are correlated, and the correlation, called "Intracluster correlation coefficient" is always positive:

$$\rho = \frac{\tau_{00}}{\sigma^2 + \tau_{00}}.$$

- Note that $\mathrm{Var}(Y_{ij}|X_{ij}) = \sigma^2 + \tau_{00}$. That is, the variance of data is decomposed to the two variance components - Level 1 variance component σ^2 and Level 2 variance component τ_{00}.
- Since u_{0i} are treated as random variables (random effect), the model is called a *mixed effect model* (mixture of the fixed effect β_1 and random effect).
- u_{0i} can be treated as fixed effects (eg. randomized block design).

- Advantages of Random effect over Fixed effects
  - Small number of parameters and so more efficient.
  - The results can be extended populationwidely.
  - Easy to incorporate complicated hierarchical structures.
  - Asymptotically valid.
- Disadvantage of Random effect over Fixed effects
  - Results may not be valid when the distribution of random effects is
    misspecified.
  - Computation are demanding for complicated hierarchial structures.

Random coefficient model

- The effects of X to Y vary across the clusters.
- Model

  (a) $Y_{ij} = \beta_{0i} + \beta_{1i} X_{ij} + e_{ij}$

  (b) $\beta_{0i} = \beta_0 + u_{0i}$

  (c) $\beta_{1i} = \beta_1 + u_{1i}$

  (d) $e_{ij} \sim N(0, \sigma^2), e_{ij} \perp (u_{0i}, u_{1i})$ and

$$\begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix} \right)$$

Estimation method

- Marginally, the model is still linear, but the errors are correlated.
- So, we can use general least square method for the fixed effects, which turns out to be the MLE.
- For variance components, several methods such as ML, REML and MINVQUE are available (in SAS Proc Mixed).
- Estimation of the fixed effects are valid asymptotically even when the underlying distribution is not normal as long as the correlation structure is correctly specified.
- However, the estimation of the variance components may not be valid. MINVQUE is robust for distribution assumption since it is a method of moment estimator.

Prediction of random effects

- Empirical Bayes approach
- First calculate

$$E(\beta_{0i}|\textbf{Data, fixed effects} \text{ and } \textbf{variance components})$$

- And replace the fixed effect and variance component by their estimators.
- It turns out that β^*_{0i}, the predicted values of β_{0i} is a convex combination of overall mean and cluster specific mean (i.e $\lambda\bar{Y}_{..} + (1-\lambda)\bar{Y}_{i.}$ for some $\lambda \in [0,1]$ when no covariate exists).
- This prediction is called a shrinkage estimator (from the fixed effect model point of view).
- It is well known that shrinkage estimators outperforms MLE, which is an advantage of using random effects.

Illustration with the SPC data

- Result of variance component model without covariate

Parameter	Estimate	SE
β_0	0.824	0.048
τ_{00}	0.068	0.020
σ^2	0.112	0.012

- Individual level variation of wakefulness exists.

- Result of variance component model with covariate

Parameter	Estimate	SE
β_0	0.671	0.059
β_0	0.138	0.034
τ_{00}	0.061	0.018
σ^2	0.105	0.011

- Cough is a significant risk factor for wakefulness.
- Still individual variation of wakefulness exists even after adjusting cough.

- Result of a random coefficient model

Parameter	Estimate	SE
β_0	0.671	0.059
β_0	0.138	0.034
τ_{00}	0.061	0.018
τ_{11}	0.026	0.016
τ_{01}	-0.027	0.020
σ^2	0.105	0.011

- τ_{11} is not significant. There appears to be not much evidence that coughing of a given amount bothers some children more than others.

- Prediction of random slope

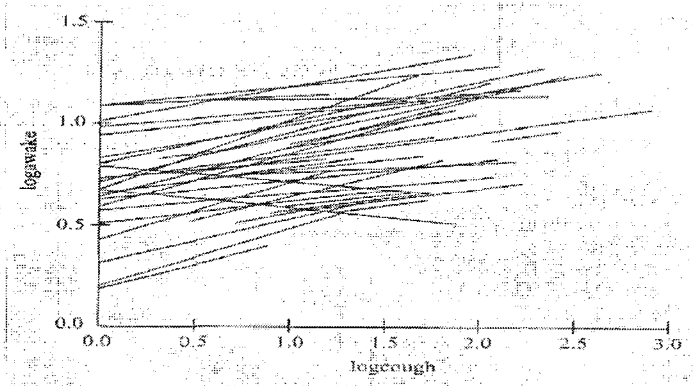

Figure 1.1 Regressions of logawake on logcough for 39 subjects.

- There appears to be two subjects with negative slopes who might be investigated further.

3. Multilevel analysis with GLM

- Multilevel model with other than normal distribution such as binary, count, survival time etc, can be done inside the framework of the GLM.
- We consider the two most popularly used such models - logistic regression model for binary data and Poisson regression for count data.

Multilevel logistic regression model

- We only present a random coefficient model.
- Model

(a) $\operatorname{logit}\Pr(Y_{ij}=1|X_{ij})=\beta_{0i}+\beta_{1i}X_{ij}$

(b) $\beta_{0i}=\beta_0+u_{0i}$

(c) $\beta_{1i}=\beta_1+u_{1i}$

(d)

$$\begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix} \right)$$

Multilevel Poisson regression model

- Model

(a) $Y_{ij} \sim Poisson(\mu_{ij})$

(b) $\log \mu_{ij}=\beta_{0i}+\beta_{1i}X_{ij}+e_{ij}$

(c) $\beta_{0i}=\beta_0+u_{0i}$

(d) $\beta_{1i}=\beta_1+u_{1i}$

(e) $e_{ij} \sim N(0,\sigma^2), e_{ij} \perp (u_{0i},u_{1i})$ and

$$\begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix} \right)$$

(*) e_{ij} term is needed for overdispersed models.

Illustration for the Multilevel Poisson regression model

- Testis cancer mortality in the European country
- Three levels: Country, regions and county
- 9 nations, 78 regions and 354 counties
- Two covariates (county level)
  - X_1: GDP per inhabitant
  - X_2: density of inhabitants per square kilometre.
- Response: the number of deaths due to testis cancer in between 1971 and 1980.
- Model: Multilevel Poisson regression model with overdispersion.

- Variance component model

Table 1: Estimation result with the PQL method

Parameter	Estimate	SE
Fixed part		
β_0	2.58	0.11
β_1	3.61	1.42
β_2	-7.22	4.71
Random part		
Level 3: nations		
$\tau_{00}^{(3)}$	0.096	0.052
Level 2: regions		
$\tau_{00}^{(2)}$	0.028	0.008
Level 1: counties		
σ^2	1.48	0.12

- Remarks
  - GDP is a significant risk factor.
  - Data is overdispersed since σ^2 is large.
  - There are significant regional variations in testis cancer mortality.
  - However, countrywide variation is not significant.

- Prediction of regional random effects

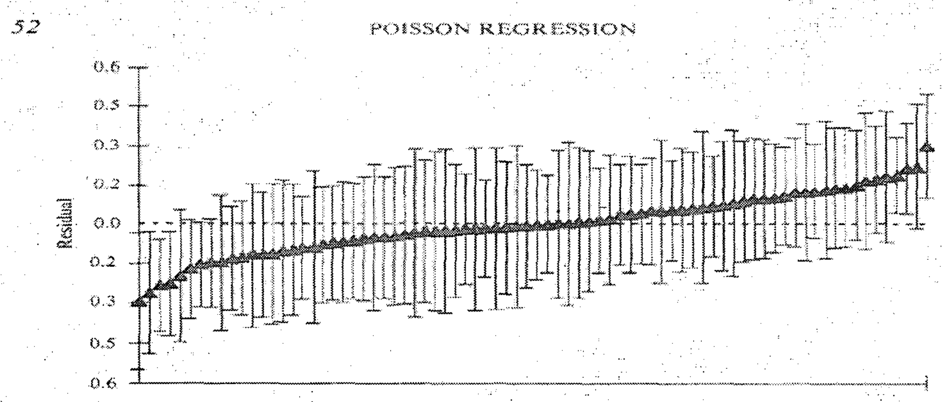

Figure 4.1 Residuals and 95% confidence intervals for the 78 regions.

- Prediction of country random effects

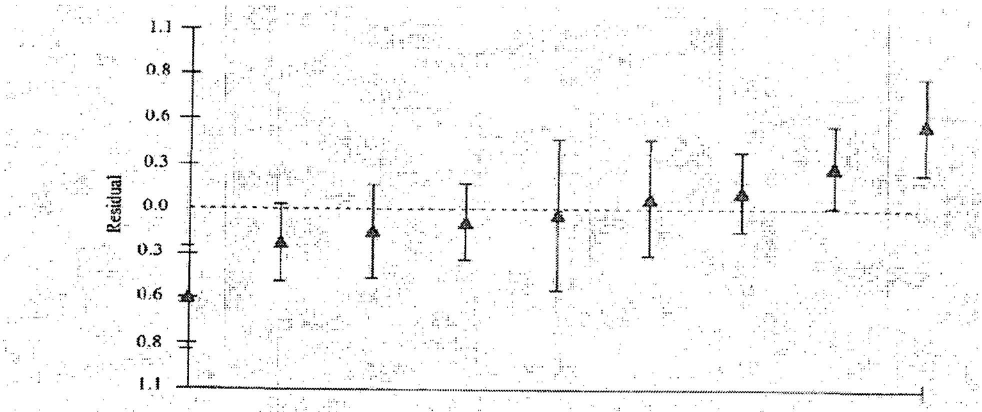

Figure 4.2 Residuals and 95% confidence intervals for the 9 countries.

- A (country level) random coefficient model only for GDP
  - Model

(a) $Y_{ij} \sim Poisson(\mu_{ij})$

(b) $\log \mu_{ij} = \beta_{0i} + \beta_{1i} X_{1ij} + \beta_2 X_{2ij} + e_{ij}$

(c) $\beta_{0i} = \beta_0 + u_{0i}$

(d) $\beta_{1i} = \beta_1 + u_{1i}$

(e) $e_{ij} \sim N(0, \sigma^2), e_{ij} \perp (u_{0i}, u_{1i})$ and

$$\begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix} \right)$$

(*) e_{ij} term is needed for overdispersed models.

- Results with PQL

Table 2: Estimation result with the PQL method

Parameter	Estimate	SE
Fixed part		
β_0	2.56	0.12
β_1	2.65	1.91
β_2	-4.74	4.92
Random part		
Level 3: nations		
$\tau_{00}^{(3)}$	0.104	0.057
$\tau_{11}^{(3)}$	8.40	10.89
$\tau_{01}^{(3)}$	-1.10	5.82
Level 2: regions		
$\tau_{00}^{(2)}$	0.028	0.008
Level 1: counties		
σ^2	1.48	0.12

- Prediction of countrylevel random coefficient

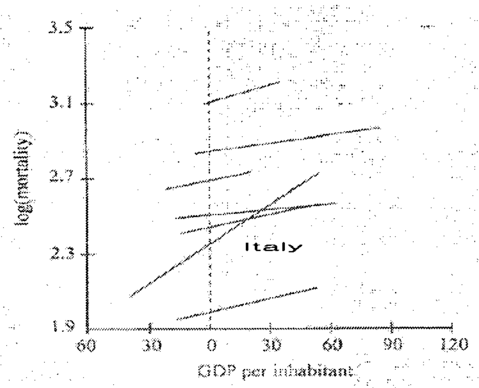

Figure 4.3 Relationship between GDP (centred) and mortality.

4. Multilevel linear model vs Multilevel GLM

- Multilevel GLM looks similar to multilevel linear models. However, there are various differences and difficulties in multilevel GLM such as
  - Interpretation of the fixed effects
  - Inferential methods
  - Choice of random effect distribution.
- We discuss differences and difficulties of multilevel linear model and multilevel GLM.

Interpretation of the fixed effect

- Consider the variance component model.
- For linear model, β_0 is the population mean of response.
- For logistic model, β_0 is not (the logit of) the population mean of Y (i.e probability).
- This is because multilevel linear models are marginally linear models while multilevel logistic models are not logistic models marginally.
- An alternative logistic model with correlated data is marginal models such as GEE. Marginal models, however, do not provide cluster level information.
- Currently, many researches for combining random effect models (subject specific model) and marginal models (population average model) have been done.

Inferential methods

- In general, the best method is to use the marginal likelihood (likelihood after integrating out random effects).
- For linear multilevel models, the marginal likelihood has closed forms and so no problem of getting MLE.
- For multilevel GLM, unfortunately, the closed form of the marginal likelihood is not available and so numerical integrations are required.
- For complicated multilevel models, there are high dimensional random effects and high dimensional numerical integrations are practically impossible.

- Alternative methods
  - Approximated marginal likelihood: PQL
  - Maximizing random effects as well as fixed effects: Hierarchical likelihood approach
  - Bayesian approach with MCMC
- Remarks
  - PQL and H-likelihood may be asymptotically inconsistent.
  - Bayesian approach may be still computationally demanding and may be inferior for small sample sizes.
- Software
  - Marginal likelihood: PROC NLMIXED (in SAS)
  - PQL: PROC GLIMMIX (SAS Macro)
  - Bayesian: WinBugs

Choice of random effect distribution

- So far, we assume that random effects are normally distributed.
- In some cases, other than normal distributions are required (eg. bimodal, skewed etc).
- For multilevel linear model, the estimators of the fixed effects are asymptotically valid even when the distribution of random effects is not normal.
- However, for multilevel GLM, misspecified random effect distributions result in biased fixed effect estimators.
- Two approaches
  - Goodness of fit for the random effect distribution
  - Nonparametric method: Mixture models.
- No practically usable software is not available yet.

5. References

- Leyland, A.H. and Goldstein, H. (2001). *Multilevel modeling of health statistics.* Wiley.
- Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized latent variable modeling.* Chapman and Hall.