

Multilevel Modeling of Non-Normal Data

Donald Hedeker, Ph.D.

Upcoming Seminar:
May 18-19, 2017, Chicago, Illinois

Multilevel Modeling of Non-Normal Data

Don Hedeker

Department of Public Health Sciences

University of Chicago

email: hedeker@uchicago.edu

Hedeker, D. (2005). Generalized linear mixed models. In B. Everitt & D. Howell (Eds.), Encyclopedia of Statistics in Behavioral Science. Wiley.

1

What are Multilevel Data?

- Data that are hierarchically structured, nested, clustered
- Data collected from units organized or observed within units at a higher level (from which data are also obtained)

<i>data collected on</i>	<i>who are clustered within</i>
students	classrooms
siblings	families
repeated observations	individuals

==> these are examples of two-level data

level 1 - (students) - measurement of primary outcome and important mediating variables

level 2 - (classrooms) - provides context or organization of level-1 units which may influence outcome; other mediating variables

2

What is Multilevel Data Analysis?

“any set of analytical procedures that involve data gathered from individuals and from the social structure in which they are embedded and are analyzed in a manner that models the multilevel structure”

L. Burstein, *Units of Analysis*, 1985, Int Ency of Educ

- analysis that *models the multilevel structure*
- recognizes influence of structure on individual outcome

<i>structure</i>	<i>may influence response from</i>
classroom	students
family	siblings
individual	repeated observations

3

Why do Multilevel Data Analysis?

- assess amount of variability due to each level (*e.g.*, family variance and individual variance)
- model level 1 outcome in terms of effects at both levels
$$\textit{individual var.} = \textit{fn}(\textit{individual var.} + \textit{family var.})$$
- assess interaction between level effects (*e.g.*, individual outcome influenced by family SES for males, not females)
- Responses are not independent - individuals within clusters share influencing factors

⇒ Multilevel analysis - another example of *Golden Rule of Statistics*: “one person’s error term is another person’s (or many persons’) career”

4

Multilevel models aka

- random-effects models
- random-coefficient models
- mixed-effects models
- hierarchical linear models

Useful for analyzing

- Clustered data
 - subjects (level-1) within clusters (level-2)
 - * e.g., clinics, hospitals, families, worksites, schools, classrooms, city wards
- Longitudinal data
 - repeated obs. (level-1) within subjects (level-2)

5

		<i>cluster variables</i>		<i>subject variables</i>		
cluster	subject	tx group	size	outcome	sex	age
1	1
	\vdots
	n_1
2	1
	\vdots
	n_2
.	1
	\vdots
	n_i
N	1
	\vdots
	n_N

$i = 1 \dots N$ clusters

$j = 1 \dots n_i$ subjects in cluster i

6

		<i>time-invariant variables</i>			<i>time-varying variables</i>	
subject	time	tx group	sex	age	outcome	dose
1	1
	⋮
	n_1
2	1
	⋮
	n_2
⋮	1
	⋮
	n_{\cdot}
N	1
	⋮
	n_N

$i = 1 \dots N$ subjects

$j = 1 \dots n_i$ timepoints for subject i

Multilevel models for categorical outcomes

- dichotomous outcomes
 - mixed-effects logistic regression
- ordinal outcomes
 - mixed-effects ordinal logistic regression
 - * proportional odds model
 - * partial or non-proportional odds model
- nominal outcomes
 - mixed-effects nominal logistic regression
- discrete or grouped time-to-event data
 - mixed-effects dichotomous or ordinal regression
 - * complementary log-log link for proportional (and non-proportional) hazards models

<http://www.ssicentral.com/supermix/Documentation/Survival.clust.pdf>

Logistic Regression Model

$$\log \left[\frac{P(Y_i = 1)}{1 - P(Y_i = 1)} \right] = \mathbf{x}'_i \boldsymbol{\beta}$$

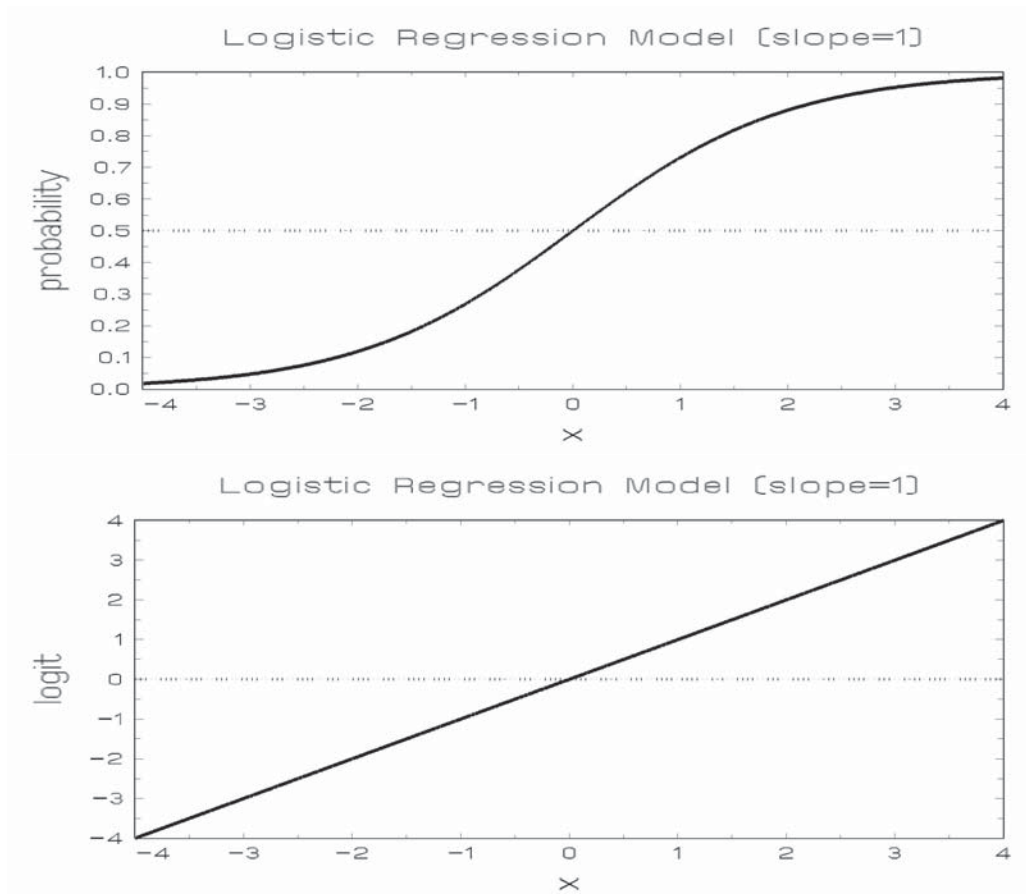
- Dichotomous outcome ($Y = 0$ absence, $Y = 1$ presence).
- Function that links probabilities to regressors is the logit (or log odds) function $\log [P/(1 - P)]$. Logit is called the link function.

The model can be written in terms of probabilities:

$$P(Y_i = 1) = \frac{1}{1 + \exp(-\mathbf{x}'_i \boldsymbol{\beta})}$$

- Model is a linear model for the logits, not for the probabilities. Logits can take on any values between negative and positive infinity, probabilities can only take on values between 0 and 1.

9



10

The model can also be written in terms of the odds:

$$\frac{P(Y_i = 1)}{1 - P(Y_i = 1)} = \exp(\mathbf{x}'_i \boldsymbol{\beta})$$

$\exp \beta$ = change in odds for Y per unit change of x

- $\beta = 0$ yields no effect on the odds
- $\beta > 0$ increases odds Y is present with increasing x
- $\beta < 0$ decreases odds Y is present with increasing x

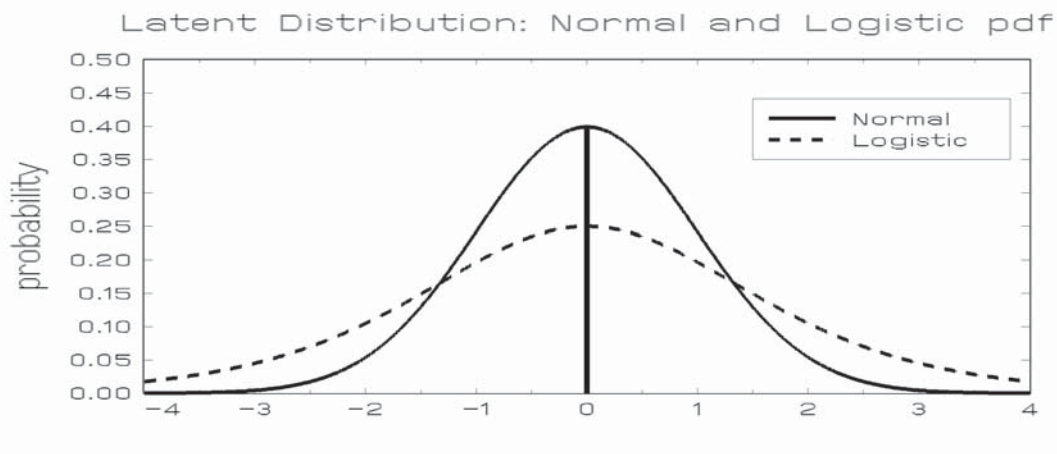
11

Dichotomous Response and Threshold Concept

Continuous y_i - an unobservable latent variable - related to dichotomous response Y_i via “threshold concept”

Response occurs ($Y_i = 1$) if $\gamma < y_i$

otherwise, a response does not occur ($Y_i = 0$)



12

The Threshold Concept in Practice

“How was your day?” (what is your satisfaction level today?)

- Satisfaction may be continuous, but we usually emit a dichotomous response:



13

Model for Latent Continuous Responses

Consider the model with p covariates for the latent response strength y_i ($i = 1, 2, \dots, N$):

$$y_i = \mathbf{x}'_i \boldsymbol{\beta} + \varepsilon_i$$

- probit: $\varepsilon_i \sim$ standard normal (mean=0, variance=1)
- logistic: $\varepsilon_i \sim$ standard logistic (mean=0, variance= $\pi^2/3$)

\Rightarrow $\boldsymbol{\beta}$ estimates from logistic regression are larger (in abs. value) than from probit regression by approximately $\sqrt{\pi^2/3} = 1.8$

Underlying latent variable

- useful way of thinking of the problem
- not an essential assumption of the model

14

Random-intercept Logistic Regression Model

Consider the model with p covariates for the response Y_{ij} for subject j ($j = 1, 2, \dots, n_i$) in cluster i ($i = 1, 2, \dots, N$):

$$\log \left[\frac{P(Y_{ij} = 1)}{1 - P(Y_{ij} = 1)} \right] = \mathbf{x}'_{ij} \boldsymbol{\beta} + v_{0i}$$

where

Y_{ij} = dichotomous response for subject j in cluster i

\mathbf{x}_{ij} = $(p + 1) \times 1$ covariate vector (includes 1 for intercept)

$\boldsymbol{\beta}$ = $(p + 1) \times 1$ vector of unknown parameters

v_{0i} = cluster effects distributed $\mathcal{NID}(0, \sigma_v^2)$
and assumed independent of \mathbf{x} variables

15

Characteristics of $v_{0i} \sim \mathcal{NID}(0, \sigma_v^2)$

- separates model from usual (fixed-effects) multiple logistic regression model
- takes on $i = 1, 2, \dots, N$ values
- assess impact of cluster i on individual outcome; represents degree of subject clustering
- common for each cluster member, but changes for each cluster
- if $v_{0i} = 0$, then cluster has no effect for cluster i
- if $v_{0i} = 0$ for all clusters, cluster structure has no impact on individual data ($\sigma_v^2 = 0$)
 - no need for multilevel approach
 - ordinary logistic regression is OK
- if subject clustering has strong effect, estimates of $v_{0i} \neq 0$ and σ_v^2 will increase from 0

16

Model for Latent Continuous Responses

Consider the model with p covariates for the $n_i \times 1$ latent response strength y_{ij} :

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + v_{0i} + \varepsilon_{ij}$$

where assuming

- $\varepsilon_{ij} \sim$ standard normal (mean 0 and $\sigma^2 = 1$) leads to multilevel probit regression
- $\varepsilon_{ij} \sim$ standard logistic (mean 0 and $\sigma^2 = \pi^2/3$) leads to multilevel logistic regression

17

Underlying latent variable

- not an essential assumption of the model
- useful for obtaining intra-class correlation (r)

$$r = \frac{\sigma_v^2}{\sigma_v^2 + \sigma^2}$$

and for design effect (d)

$$d = \frac{\sigma_v^2 + \sigma^2}{\sigma^2} = 1/(1 - r)$$

ratio of actual variance to the variance that would be obtained by simple random sampling (holding sample size constant)

18

Scaling of regression coefficients

Fixed-effects model

β estimates from logistic regression are larger (in abs. value) than from probit regression by approximately

$$\sqrt{\frac{\pi^2/3}{1}} = 1.8$$

because

- $V(y) = \sigma^2 = \pi^2/3$ for logistic
- $V(y) = \sigma^2 = 1$ for probit

Mixed-effects model

β estimates from mixed-effects model are larger (in abs. value) than from fixed-effects model by approximately

$$\sqrt{d} = \sqrt{\frac{\sigma_v^2 + \sigma^2}{\sigma^2}}$$

because

- $V(y) = \sigma_v^2 + \sigma^2$ in mixed-effects model
- $V(y) = \sigma^2$ in fixed-effects model

difference depends on size of random-effects variance σ_v^2

Within-Clusters / Between-Clusters models

Within-clusters model - level 1 ($j = 1, \dots, n_i$)

$$\begin{aligned} & \text{observed response} \\ \log \left[\frac{P(Y_{ij} = 1)}{1 - P(Y_{ij} = 1)} \right] &= b_{0i} + b_{1i} \text{Sex}_{ij} \\ & \text{latent response} \\ y_{ij} &= b_{0i} + b_{1i} \text{Sex}_{ij} + \varepsilon_{ij} \end{aligned}$$

Between-clusters model - level 2 ($i = 1, \dots, N$)

$$\begin{aligned} b_{0i} &= \beta_0 + \beta_2 \text{Grp}_i + v_{0i} \\ b_{1i} &= \beta_1 + \beta_3 \text{Grp}_i \end{aligned}$$

with $v_{0i} \sim \mathcal{NID}(0, \sigma_v^2)$ and $\varepsilon_{ij} \sim \mathcal{LID}(0, \pi^2/3)$

21

Put together,

$$\begin{aligned} \text{logit}_{ij} &= b_{0i} + b_{1i} \text{Sex}_{ij} \\ &= (\beta_0 + \beta_2 \text{Grp}_i + v_{0i}) + (\beta_1 + \beta_3 \text{Grp}_i) \text{Sex}_{ij} \\ &= \beta_0 + \beta_1 \text{Sex}_{ij} + \beta_2 \text{Grp}_i + \beta_3 (\text{Grp}_i \times \text{Sex}_{ij}) + v_{0i} \end{aligned}$$

- $\beta_0 = \text{logit when } \text{Sex} = \text{Grp} = 0$
- $\beta_1 = \text{Sex effect when } \text{Grp} = 0$
- $\beta_2 = \text{Grp effect when } \text{Sex} = 0$
- $\beta_3 = \text{difference between Sex effect for Grp} = 1 \text{ vs } \text{Grp} = 0;$
or difference between Grp effect for Sex = 1 vs Sex = 0

\Rightarrow coding of variables very important for correct interpretation.
Also, these are controlling for cluster effect (“cluster-specific” effects)

22

Effects of a School-based Intervention

The Television School and Family Smoking Prevention and Cessation Project (Flay, *et al.*, 1988); a subsample:

- *sample* - 1600 7th-graders - 135 classes - 28 schools
 - 1 to 13 classes per school, 2 to 28 students per class
- *outcome* - knowledge of the effects of tobacco use
- *timing* - students tested at pre and post-intervention
- *design* - schools exposed to
 - a social-resistance classroom curriculum (CC)
 - a media (television) intervention (TV)
 - CC combined with TV
 - a no-treatment control group

23

Main question of interest:

- Influence of the intervention on the tobacco health knowledge scores (THKS) ?

Challenges in the analysis:

- outcome variable (THKS) is number correct of 7 items
- controlling for intra-school and intra-class variability
- potential explanatory variables are at different levels

24

Tobacco and Health Knowledge Scale
 Post-Intervention Scores ≥ 3 (out of 7)
 Subgroup Descriptive Statistics

	CC = no		CC = yes	
	TV=no	TV=yes	TV=no	TV=yes
n	421	416	380	383
proportions	.416	.483	.632	.603
odds	.711	.935	1.714	1.520
logits	-.341	-.067	.539	.419

Within-Clusters / Between-Clusters components

Within-clusters model - level 1 ($j = 1, \dots, n_i$ subjects)

$$\text{logit}_{ij} = b_{0i}$$

Between-clusters model - level 2 ($i = 1, \dots, N$ clusters)

$$b_{0i} = \beta_0 + \beta_1 CC_i + \beta_2 TV_i + \beta_3 (CC_i \times TV_i) + v_{0i}$$

$$v_{0i} \sim \mathcal{NID}(0, \sigma_v^2)$$

β_0 = THKS logit for CC=no TV=no subgroup

β_1 = logit diff. between CC=yes vs CC=no (for TV=no)

$$b_{0i} = \beta_0 + (\beta_1 + \beta_3 TV_i) CC_i + \beta_2 TV_i + v_{0i}$$

β_2 = logit diff. between TV=yes vs TV=no (for CC=no)

$$b_{0i} = \beta_0 + (\beta_2 + \beta_3 CC_i) TV_i + \beta_1 CC_i + v_{0i}$$

β_3 = difference in logit attributable to interaction

v_{0i} = random cluster deviation

note: interpretation depends on coding of variables, and β s are adjusted for the cluster effects (cluster-specific effects)

27

3-level model

Within-classrooms (and schools) model - level 1
($k = 1, \dots, n_{ij}$ students)

$$\text{logit}_{ijk} = b_{0ij}$$

Between-classrooms (within-schools) model - level 2
($j = 1, \dots, n_i$ classrooms)

$$b_{0ij} = b_{0i} + v_{0ij}$$

Between-schools model - level 3 ($i = 1, \dots, N$ schools)

$$b_{0i} = \beta_0 + \beta_1 CC_i + \beta_2 TV_i + \beta_3 (CC_i \times TV_i) + v_{0i}$$

$$v_{0ij} \sim \mathcal{NID}(0, \sigma_{v(2)}^2) \quad \text{and} \quad v_{0i} \sim \mathcal{NID}(0, \sigma_{v(3)}^2)$$

28

β_0 = THKS logit for CC=no TV=no subgroup

β_1 = logit diff. between CC=yes vs CC=no (for TV=no)

β_2 = logit diff. between TV=yes vs TV=no (for CC=no)

β_3 = difference in logit attributable to interaction

v_{0ij} = random classroom deviation

v_{0i} = random school deviation

Stata for multilevel analysis of dichotomous outcomes: melogit (version 13 and thereafter)

- Multiple levels of nesting, crossed random effects
- Full likelihood estimation using numerical quadrature for integration over the random effects
 - non-adaptive, mode/curvature adaptive, mean/variance adaptive (default except for crossed random effects)
 - 7 points per dimension are the default; more points provides greater accuracy, but also more computation time
- Laplace approximation (default for crossed random effects models)
 - same as mode/curvature adaptive with one point
 - can produce biased estimates, especially as the ICC is high and numbers of clusters and/or subjects is small