STATISTICAL HORIZONS

Multilevel Modeling of Categorical Outcomes

Donald Hedeker, Ph.D.

Upcoming Seminar:

December 16-18, 2021, Remote Seminar

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.

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)

data collected on	*who are clustered within*
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

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

structure	*may influence response from*
classroom	students
family	siblings
individual	repeated observations

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

$$individual\ var. = fn(individual\ var. + family\ var.)$$

- 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"

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)

		cluster variables		*subject variables*		
cluster	subject	tx group	size	outcome	sex	age
1	1	.	.	.	.	.
	$\vdots$	.	.	.	.	.
	n_1	.	.	.	.	.
2	1	.	.	.	.	.
	$\vdots$	.	.	.	.	.
	n_2	.	.	.	.	.
.	1	.	.	.	.	.
	$\vdots$	.	.	.	.	.
	$n_.$	.	.	.	.	.
N	1	.	.	.	.	.
	$\vdots$	.	.	.	.	.
	n_N	.	.	.	.	.

$i = 1 \ldots N$ clusters
$j = 1 \ldots n_i$ subjects in cluster i

		time-invariant	*variables*		*time-varying*	*variables*
subject	time	tx group	sex	age	outcome	dose
1	1	.	.	.	.	.
	⋮	.	.	.	.	.
	n_1	.	.	.	.	.
2	1	.	.	.	.	.
	⋮	.	.	.	.	.
	n_2	.	.	.	.	.
.	1	.	.	.	.	.
	⋮	.	.	.	.	.
	$n_.$	.	.	.	.	.
N	1	.	.	.	.	.
	⋮	.	.	.	.	.
	n_N	.	.	.	.	.

$i = 1 \ldots N$ subjects
$j = 1 \ldots n_i$ 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

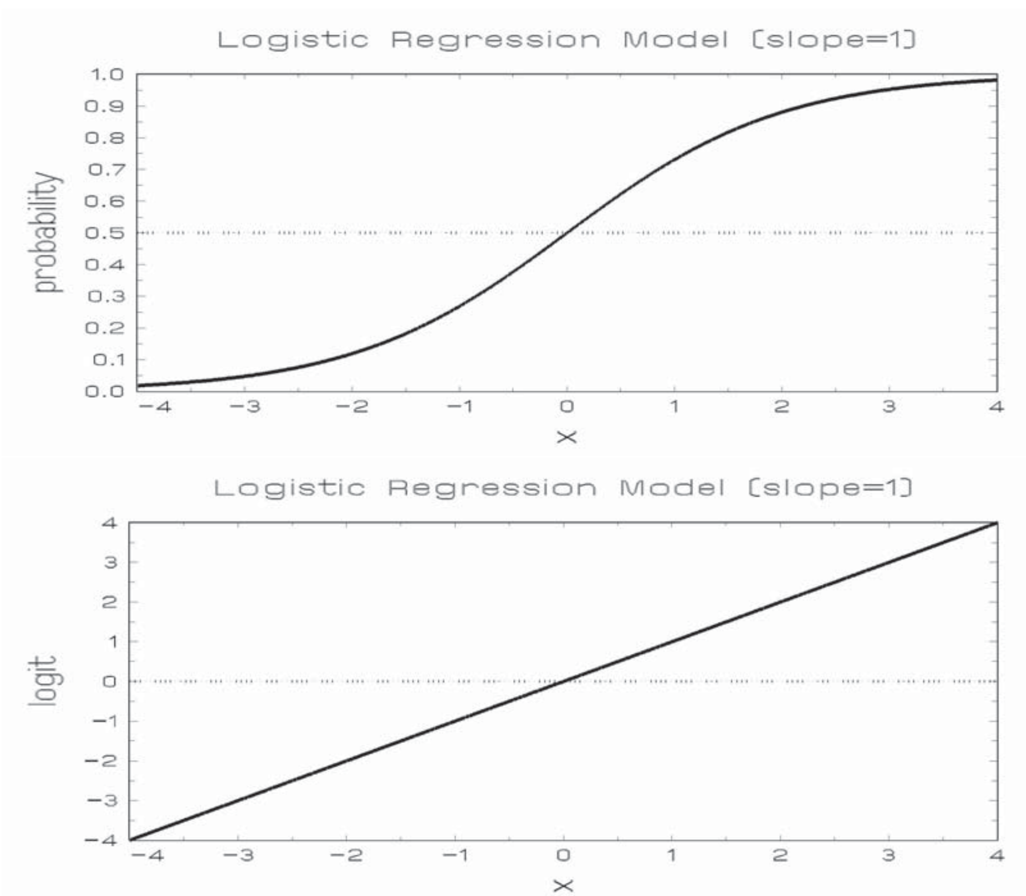
$$\log\left[\frac{P(Y_i = 1)}{1 - P(Y_i = 1)}\right] = \boldsymbol{x}_i'\boldsymbol{\beta}$$

- 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:

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

- 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.



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

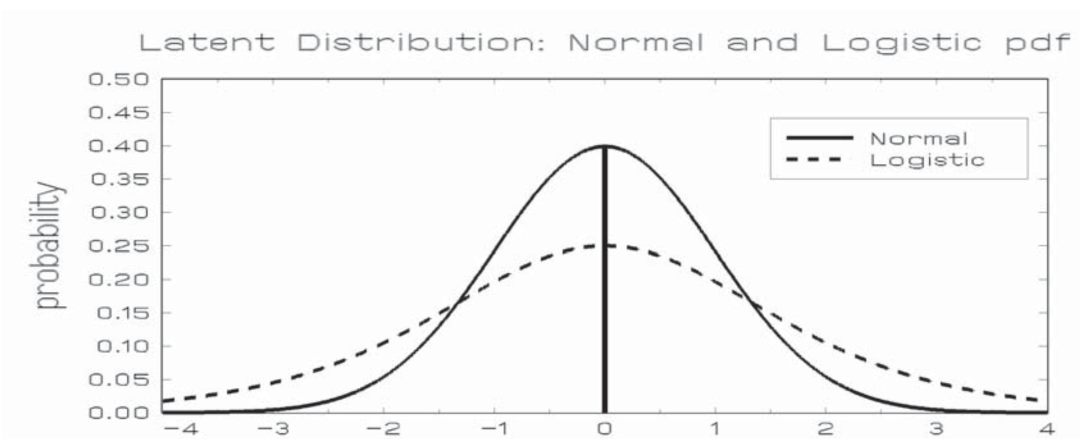
$$\frac{P(Y_i = 1)}{1 - P(Y_i = 1)} = \exp(\boldsymbol{x}_i'\boldsymbol{\beta})$$

$\exp\beta$ = 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

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 ($Y_i = 0$)



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:

Model for Latent Continuous Responses

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

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

- 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

Random-intercept Logistic Regression Model

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

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

where

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

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

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

υ_{0i} = cluster effects distributed $\mathcal{NID}(0, \sigma^2_\upsilon)$ and assumed independent of $\boldsymbol{x}$ variables

Characteristics of $\upsilon_{0i} \sim \mathcal{NID}(0, \sigma^2_\upsilon)$

- separates model from ususal (fixed-effects) multiple logistic regression model
- takes on $i = 1, 2, \ldots, N$ 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 $\upsilon_{0i} = 0$, then cluster has no effect for cluster i
- if $\upsilon_{0i} = 0$ for all clusters, cluster structure has no impact on individual data ($\sigma^2_\upsilon = 0$)
  - no need for multilevel approach
  - ordinary logistic regression is OK
- if subject clustering has strong effect, estimates of $\upsilon_{0i} \neq 0$ and σ^2_υ will increase from 0

Model for Latent Continuous Responses

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

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

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

Underlying latent variable

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

$$r = \frac{\sigma_\upsilon^2}{\sigma_\upsilon^2 + \sigma^2}$$

  and for design effect (d)

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

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

Scaling of regression coefficients

Fixed-effects model
$\boldsymbol{\beta}$ 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
$\boldsymbol{\beta}$ estimates from mixed-effects model are larger (in abs. value) than from fixed-effects model by approximately

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

because

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

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

Within-Clusters / Between-Clusters models

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

observed response

$$\log \left[\frac{P(Y_{ij} = 1)}{1 - P(Y_{ij} = 1)} \right] = b_{0i} + b_{1i} \, Sex_{ij}$$

latent response

$$y_{ij} = b_{0i} + b_{1i} \, Sex_{ij} + \varepsilon_{ij}$$

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

$$b_{0i} = \beta_0 + \beta_2 \, Grp_i + \upsilon_{0i}$$
$$b_{1i} = \beta_1 + \beta_3 \, Grp_i$$

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

Put together,

$$\begin{aligned} logit_{ij} &= b_{0i} + b_{1i} \, Sex_{ij} \\ &= (\beta_0 + \beta_2 \, Grp_i + \upsilon_{0i}) + (\beta_1 + \beta_3 \, Grp_i) Sex_{ij} \\ &= \beta_0 + \beta_1 \, Sex_{ij} + \beta_2 \, Grp_i + \beta_3 \, (Grp_i \times Sex_{ij}) + \upsilon_{0i} \end{aligned}$$

- β_0 = logit when $Sex = Grp = 0$
- β_1 = Sex effect when $Grp = 0$
- β_2 = Grp effect when $Sex = 0$
- β_3 = difference between Sex effect for $Grp = 1$ vs $Grp = 0$; or difference between Grp effect for $Sex = 1$ vs $Sex = 0$

⇒ coding of variables very important for correct interpretation. Also, these are controlling for cluster effect ("cluster-specific" effects)

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

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

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 $\quad (j = 1, \ldots, n_i$ subjects)

$$logit_{ij} = b_{0i}$$

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

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

$$\upsilon_{0i} \sim \mathcal{NID}(0, \sigma_\upsilon^2)$$

$$\beta_0 \;=\; \text{THKS logit for CC=no TV=no subgroup}$$

$$\beta_1 \;=\; \text{logit diff. between CC=yes vs CC=no (for TV=no)}$$

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

$$\beta_2 \;=\; \text{logit diff. between TV=yes vs TV=no (for CC=no)}$$

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

$$\beta_3 \;=\; \text{difference in logit attributable to interaction}$$

$$v_{0i} \;=\; \text{random cluster deviation}$$

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

3-level model

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

$$logit_{ijk} = b_{0ij}$$

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

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

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

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

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

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

υ_{0ij} = random classroom deviation

υ_{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